TOWARDS TRUSTWORTHY DECISION MAKING IN VISUAL ANALYTICS

WENKAI HAN

PH.D. THESIS

November 2022

Supervisor: Hans-Jörg Schulz

AARHUS UNIVERSITY
DEPARTMENT OF COMPUTER SCIENCE

TOWARDS TRUSTWORTHY DECISION MAKING IN VISUAL ANALYTICS

WENKAI HAN

Ph.D. Thesis

Department of Computer Science
Faculty of Natural Sciences
Aarhus University

November 2022

A disciple of Confucius, Zengzi, remarked,
"I daily examine myself on three counts:
Have I exhausted all efforts in the duties entrusted by others?
Have I been sincere and trustworthy with my contacts?
Have I practiced what I profess in my teaching?"

— *The Analects*

曾子曰:
“吾日三省吾身:
为人谋而不忠乎?
与朋友交而不信乎?
传不习乎?”

—《论语》

ABSTRACT

Combining the power of data analysis and interactive visualization, Visual Analytics (VA) connects the distinct capabilities of humans and machines. This further enables the human users to effectively understand data, conduct analyses, and produce insights. However, to produce truthful insights through VA, it remains an arduous challenge to support users to make trustworthy decisions in each step of VA processes. For example – How can one trust their choice of clustering algorithm out of the dozens available in modern VA tools? How can one trust their focus on one particular region in a visualization and not another? How can one trust their recollection of important data points on which their analytical decisions are based? Yet, there is still little research on what trustworthy decision making (DM) in VA encompasses and how to support it. Facing this research challenge, I explore the topic of trustworthy DM in VA from three perspectives in this thesis – theoretical foundations, a generic method, and alternative techniques.

First, I present theoretical foundations on DM and trust in VA. To this end, I introduce related DM theories and inspect the concept of trust in each element of VA processes. Second, I present a generic method that supports trustworthy DM in VA through guidance. To this end, I dissect VA guidance as decision support and propose a step-by-step guidance method for supporting trustworthy DM in VA. Third, I present empirical studies on two alternative techniques that can be used to support trustworthy DM in VA – – vibrotactile guidance and sketchy rendering. To this end, I outline the corresponding design space for each of the two techniques, summarize the insights from the empirical studies, and discuss their utilities for aiding trustworthy DM in VA.

In summary, this thesis makes contributions in a threefold manner: the theoretical foundations establish and delineate the concept of trustworthy DM in VA, the generic method reframes guidance as a widely applicable approach to support trustworthy DM in VA, and the two techniques provide practical insights for supporting trustworthy DM in VA with empirical studies. These contributions connects research in VA, DM, and trust, while providing new perspectives on VA guidance.

RESUMÉ

I kombinationen af data analyse og interaktive visualiseringer forbinder Visual Analytics VA mennesker og maskiners distinkte evner. Dette gør det muligt for brugere effektivt at kunne forstå, udføre analyser på og producere indblik i data. For at kunne producere sandfærdige indblik gennem VA er det dog stadig en svær udfordring at understøtte brugere i at gøre troværdige valg gennem hvert enkelt trin i VA processer. Eksempelvis - Hvordan kan man have tillid til sin beslutning om at fokusere på et bestemt område i en visualisering og ikke en anden? Hvordan kan man have tillid til sit valg af en clustering algoritme ud af de mange som er tilgængelige i moderne VA værktøjer? Hvordan kan man have tillid til sine erindringer om vigtige datapunkter, som ens analytiske beslutninger er baseret på? Der er stadig begrænsede mængder af forskning på hvad troværdig decision making (DM) i VA omfatter, og hvordan det kan understøttes. I denne afhandling udforsker jeg emnet troværdig DM i VA fra tre perspektiver - teoretiske grundlag, en generisk metode, og to forskellige teknikker.

Først præsenterer jeg et teoretisk grundlag for DM og tillid i VA. I dette øjemed introducerer jeg relaterede DM teorier og undersøger konceptet tillid igennem hver del af VA processer. Dernæst præsenterer jeg en generisk metode til at understøtte troværdig DM i VA gennem guidance. I dette øjemed udforsker jeg VA guidance som en form for decision support og fremsætter en trin for trin guidance metode til at understøtte troværdig DM i VA. Til sidst præsenterer jeg empiriske studier på to forskellige teknikker til at understøtte troværdig DM i VA - vibrotactile guidance og sketchy rendering. I dette øjemed optegner jeg det givne design space for hver af de to teknikker, opsummerer indblik fra de empiriske undersøgelser, og diskuterer deres anvendelighed til at støtte troværdig DM i VA.

Denne afhandling bidrager videnskabeligt på tre måder: De teoretiske grundlag etablerer og skildrer konceptet troværdig DM i VA, den generiske metode gentænker guidance som en bredt anvendelig tilgang til at understøtte troværdig DM i VA, og de to teknikker giver gennem empiriske studier praktiske indblik til at understøtte troværdig DM i VA. Disse bidrag forbinder forskning i VA, DM og troværdighed, og giver nye perspektiver på VA guidance.

ACKNOWLEDGMENTS

The three years during my PhD at Aarhus University have been challenging but rewarding. I have grown a lot both academically and personally. I would like to thank all the people who have been involved in the process.

I would first like to thank my supervisor, Hans–Jörg Schulz, who introduced me to the research world of VA and has always challenged me to go outside of my comfort zone. I would not be able to finish this thesis without your guidance.

I also would like to thank all my colleagues at Aarhus University, who made me feel at home while being thousands of kilometers away from the city I grew up in. In particular, thank you to Marius & Jana, who walked through my PhD together in the same office & created a safe space to support one another. Nathalie, thank you for the support every time I dropped by your office and when we went out together. Susanne and Henrik, thank you for supporting me like friends and giving me great advice on how to navigate myself through the PhD.

Thank you to my friends for their support over the years. Thank you to 赵欣然, who has always motivated through the hardest times in my life and is not afraid to have the hardest but necessary conversations. Ida and Martin, who helped me feel like I belong in a new country and motivated me to take care of my health. 潘耕, 张正卓, 蔡颖佳, who created cultural comfort zone for me feel safe and talk about our shared struggles as Chinese expats living in Denmark. Everyone at Run for Friendship and LGBT+HUSET, who gave me a community to meet new friends and supported me through the extreme isolation during COVID.

Last but not least, I would like to dedicate this thesis to my family, especially my late mother, 李敏, who gave me life and taught me to be a hardworking, honest, and caring person. I will forever miss you, and I know you will be with me wherever I go.

韩文锴 *(Wenkai Han)*,

Copenhagen, November 15, 2022.

PUBLICATIONS & MANUSCRIPTS

This thesis proceeds in two parts: Part I provides an overview of my research contributions throughout the three years in my PhD, and Part II contains the publications and a manuscript listed below that support these contributions.

Paper A

W. **Han**. "Making and Trusting Decisions in Visual Analytics." In: *IEEE VIS 2021 Workshop on TRust and EXpertise in Visual Analytics (TREX)*. IEEE, 2021, pp. 14–19. DOI: 10.1109/TREX53765.2021.00008.

Paper B

W. **Han** and H.-J. Schulz. "Beyond Trust Building — Calibrating Trust in Visual Analytics." In: *IEEE VIS 2020 Workshop on TRust and EXpertise in Visual Analytics (TREX)*. IEEE, 2020, pp. 9–15. DOI: 10.1109/TREX51495.2020.00006.

Paper C

W. **Han** and H.-J. Schulz. "Designing and Providing Visual Analytics Guidance through Decision Support." In: *Information Visualization (accepted with minor revision)* (2022).

Paper D

W. **Han** and H.-J. Schulz. "Exploring Vibrotactile Cues for Interactive Guidance in Data Visualization." In: *Proceedings of the 13th International Symposium on Visual Information Communication and Interaction*. Association for Computing Machinery, 2020. DOI: 10.1145/3430036.3430042.

Paper E

M. R. E. Larsen, W. **Han**, and H.-J. Schulz. "Sketchy Rendering to Aid the Recollection of Regular Visualizations." In: *22nd Eurographics Conference on Visualization*. The Eurographics Association, 2020. DOI: 10.2312/evs.20201062.

CONTENTS

ACRONYMS

VA	Visual Analytics
DM	Decision Making
InfoVis	Information Visualization
HCI	Human-Computer Interaction
MCDA	Multi-Criteria Decision Analysis
DSS	Decision Support Systems
IS	Information Systems

Part I

OVERVIEW

1 INTRODUCTION

Since Alan Turing envisioned the "universal Turing machine" in 1936, computer scientists have been relentlessly developing new computer architectures and algorithms that undertake data computations with higher accuracy and efficiency. These developments have led to the world we currently live in, where enormous amounts of data are being produced, collected, and analyzed every millisecond. From personal data such as biometrics through our wearables [10] and posts that we make on social media [7], to public data collected by sensors installed for road traffic [149] and climate analysis [254], these ubiquitous data are also constantly being processed to inform all kinds of decisions made in our lives – they can be as small as planning personal workouts and choosing commercial products online, but can also be as influential as designing public transportation and combating climate change.

Coupling the data processing capacity of machines with the knowledge and experience of human experts, Visual Analytics (VA) enables the integration of the unique capabilities between the two through interactive visualization [257]. Such integration can bring out the "best of both worlds" from humans and machines, appropriately utilizing the computing power with corresponding processing algorithms, intuitively enabling analysts to identify patterns with their expertise, and efficiently producing well-informed and critical insights that otherwise would be impossible to discover [134]. However, these benefits do not come without challenges. On the one hand, the trustworthiness of data and computation has raised major concerns in both Computer Science research and public discourse – when already biased data sources [91] feed into potentially overfitted computational models [36], the produced discriminatory results will solidify and magnify existing human biases from racial discrimination in judiciary systems [264] to unfairness based on school districts and social classes in education systems [176]. On the other hand, steering the analytical processes, human actors are ultimately responsible for weaving the data and computation together – from choosing and cleaning data sources, to constructing the processing pipelines and algorithms – every single human decision along the way heavily influences the analytical results [157]. These analytical decisions in turn inform external decisions that could change the lives of millions, yet are subject to various and common pitfalls in human decision making (DM).

Abundant research has shown how human decisions can be untrustworthy in a wide range of manners. From different perceptual errors [113, 266] and insufficient or ill-informed knowledge [222], to the impacts of various cognitive biases [66, 74] and contextual factors [24] – potential pitfalls in every element of human decisions can lead the course of the analytical process towards a false end. A recent crowdsourced study with staggering 179 authors shows that scientists experienced in data analysis can draw drastically different, sometimes opposing conclusions, even when they are testing the same hypotheses with an identical dataset, and detailed justification is required for their analytical decisions [232]. This is further exacerbated by the growing complexities in VA systems – including the analyzed data [251], implemented VA methods [216], and interactive visualizations [122] – all of which put even higher requirements on the human users to comprehensively understand the system and carefully make their analytical decisions with a sufficient and well-informed knowledge base. Therefore, to ensure the trustworthiness of the influential decisions that the analytical results inform, it is vital to design VA systems in ways that support the trustworthiness of human decisions during the analytical processes.

Against such a background, this thesis aims to explore the concept of trustworthy DM and provide insights towards supporting the trustworthiness of human decisions in VA. In the rest of this chapter, I first lay out the research challenges based on existing research and motivate the research questions I aim to answer in this thesis, then summarize the research approaches taken in this thesis to address these research challenges and questions, and finally describe the outline for the rest of this thesis and my contributions in each publication and manuscript. An overview of the research questions, approaches, and papers is presented in Figure 1.1.

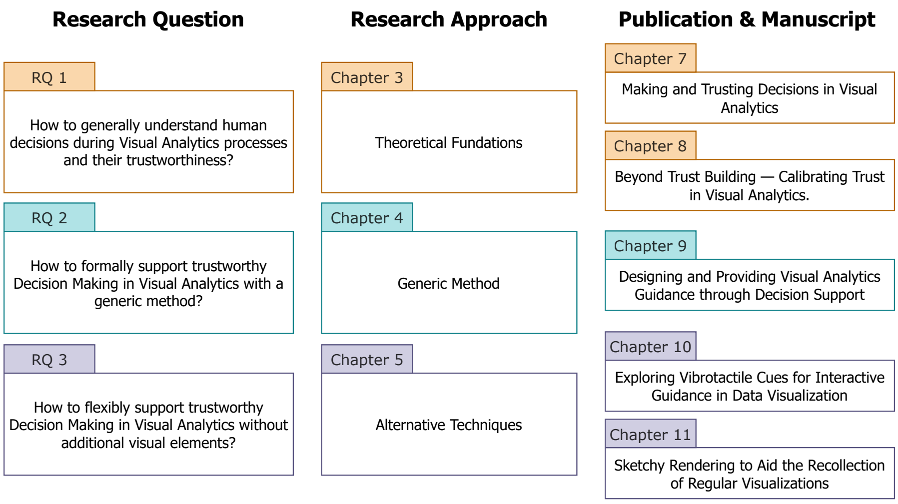


Figure 1.1: The structure of this thesis with the corresponding research question, research approach, and publication & manuscript.

1.1 RESEARCH CHALLENGES AND QUESTIONS

Both humans and machines contribute to the decisions in VA with their unique capabilities. To effectively leverage these capabilities for trustworthy DM, VA systems need to not only properly utilize the computing power of the machines to process the data, but also support human users to truthfully integrate relevant knowledge, experience, and expertise into their analytical decisions. However, previous VA research puts noticeably more emphasis on the external and final decisions that VA systems lead to than the decisions along the analytical processes. In particular, DM is often omitted in the task and interaction taxonomies [67], making it hard, if not impossible, to identify the human decisions made in the VA processes, not to mention to analyze and support these decisions. A potential reason for this is that different levels of granularity of DM are often entangled with each other – ranging from the final decisions that the analyses aim to aid [23], to a single action of zooming and panning during the analytical processes [97] – blurring the concept of DM in VA.

There also exist complex trust dynamics between the two entities in VA – machines and humans. Most of previous research has focused on the trustworthiness of machines – a wide range of research from VA [59, 221], Information Visualization (InfoVis) [42, 172, 285], HCI [55, 94, 114, 286], and automation [112, 150] has discussed how trust propagates through different elements in machines and how human users perceive the trustworthiness of machines. Meanwhile, the trustworthiness of human users is less studied in VA. VA and InfoVis research does exist in specific techniques to support users in their analytical processes, such as providing relevant information [131], mitigating human biases [270], and verifying hypotheses [130]. However, these techniques are often not explicitly discussed in the context of the trustworthiness of human decisions, and an overall understanding of what exactly the trustworthiness of human decisions means and encompasses is still lacked. Especially with the rising concept of mixed-initiative in VA research where the machines and human users collaborate and make decisions in a co-adaptive manner [246], understanding and supporting the trustworthiness of human decisions should be treated as important as that of machines' to fully enable the co-adaptive dynamics.

To support trustworthy DM in VA, it is therefore essential to first clearly understand what trustworthy DM in VA means. Therefore, this thesis aims to gain such understanding by examining and connecting previous research in VA, DM, and trust. This brings forward the first overarching research question in this thesis:

RQ1 *How to generally understand human decisions during VA processes and their trustworthiness?*

After putting DM and trust in the context of VA, the next challenge is then how to utilize these concepts and theories to support trustworthy DM in VA. As relevant research in VA aimed primarily at addressing specific applications [96, 130, 131], it is often difficult to apply these scenario-specific studies more generally. Therefore, providing a generic method to explicitly support DM in VA is important to achieve this goal in a wide range of VA scenarios. However, as DM tasks are often overlooked in previous VA research [67], how to identify the analytical decisions in VA and how to integrate related research to support these decisions is still largely an unresolved research challenge. Therefore, this thesis aims to construct a generic method that structurally identifies human decisions in their analytical processes and formally supports the trustworthiness of these decisions through integrating related research in VA, DM, and trust. This helps to answer the second research question of this thesis:

RQ2 *How to formally support trustworthy DM in VA with a generic method?*

Another research challenge that comes with adding decision support to VA systems is the additional complexity it might bring. As VA relies mainly on the visual channel to communicate complex data and methods, many VA systems are often already visually loaded and put a strain on the cognitive ability of users [161, 188], which has a negative impact on the quality and trustworthiness of human decisions [62]. In these visually complex VA systems, it can be difficult to find available visual space and encoding, and adding even more visual elements to support DM could be counterproductive in these scenarios. Furthermore, many decisions in VA are dynamic and flexible, such as zooming in and out between different data areas, experimenting with divergent data processing pipelines, and changing various specifications of visualizations. These exploratory decisions are subject to change as users gradually proceed with their analyses, while supporting such decisions with elaborate and formalized DM models could also bring in unnecessary complexity and can limit users from freely exploring different paths in their analyses. Therefore, this thesis aims to develop alternative techniques that could flexibly support trustworthy DM in visually complex VA systems without adding visual elements, which helps answer the third and final research question of this thesis:

RQ3 *How to flexibly support trustworthy DM in VA without additional visual elements?*

1.2 RESEARCH APPROACH

To address the research questions above, this thesis is conceptually structured in a threefold manner – theoretical foundations, generic method, and alternative techniques – with the theoretical analysis providing the foundations for building the method and techniques, the generic method as the intermediate medium directly applying the theories while considering practical applications, and the techniques informing the theory and method with empirical insights. These three general approaches also correspond to the three research questions that this thesis aims to answer, as illustrated in Figure 1.1.

First, to support trustworthy decisions, an understanding of what trust and DM mean in VA is needed. To this end, I first reviewed theories from related research areas and analyzed them in the context of VA. For DM, I analyzed how different DM theories in information discovery, decision strategies, reasoning and rationality, and decision analysis relate to VA and facilitate trustworthy human decisions. For trust, I reviewed the different concepts in related research, such as trustworthiness and different levels of trust, and inspected the possible factors related to trust in the human-machine analytical process by going through each of its components based on previous VA research.

Second, to provide a widely applicable solution to formally support trustworthy DM in VA, a generic method was developed through connecting the concept of guidance with related research in VA, DM, and HCI. On the one hand, the guidance method builds upon the theoretical foundations – it was framed around the concept of decision points, analyzed scenarios where guidance is needed through DM processes, and constructed the core of the guidance generation with MCDA from DM research. On the other hand, the guidance method informs the development of practical applications and techniques – a set of worksheets was produced based on a series of expert workshops to facilitate the use of the generic method, and an exemplary implementation was built to showcase how to apply the method.

Finally, alternative techniques were built to support DM in VA without adding visual elements, while producing practical insights with empirical evidence. The two techniques – vibrotactile guidance and sketchy rendering – were built under their corresponding design space produced through reviewing related research. For vibrotactile guidance, a list of widely available devices was also analyzed against the parameters in the design space. Subsequently, the prototypes in both studies were built under the technical constraints of the available devices and refined with pilot studies. Finally, both quantitative (user performance and rating) and qualitative approaches (semi-structured interview) were combined during the user studies of the developed prototypes. The different approaches used in the user studies come with their own unique benefits and complement one another – user performance offers objective metrics such as speed and accuracy to evaluate the different designs in each technique and uncover how these designs influence user interactions unconsciously, user rating complements the metrics in user performance by quantitatively capturing subjective user experience, while semi-structured interview provides the opportunities for users to further elaborate on their experiences and help to interpret the quantitative results.

1.3 THESIS OUTLINE & CONTRIBUTIONS

Overall, this thesis is structured with two parts – an overview in Part I and the supporting papers in Part II.

In Part I, I provide an overview of the contributions in this thesis:

To introduce the background which my thesis builds upon, I present the related work among the interdisciplinary fields of VA, DM, trust, and guidance in Chapter 2. Building on this background, the contributions of this thesis are then presented in a threefold manner based on the research questions, approaches, and papers:

First, to clearly delineate the important concept of trustworthy DM in VA, I present its theoretical foundations through dissecting the concept of DM in VA and the trustworthiness of human decisions in VA in Chapter 3. This chapter also serves as a supplement to the research background in Chapter 2, which is informed by the publications in Chapter 7 and Chapter 8.

Second, to provide a widely applicable approach for supporting trustworthy DM in VA, I present the generic guidance method through decision support, discuss how it foster trustworthy DM, and relate this method to previous guidance research in Chapter 4. This chapter is informed by the accepted manuscript in Chapter 9.

Third, to support trustworthy DM in VA without adding significant mental burden, I present two alternative techniques – vibrotactile guidance and sketchy rendering, and discuss their utilities based on empirical user studies in Chapter 5. This chapter is informed by the publications in Chapter 10 and Chapter 11.

Finally, I highlight potential directions of future work and summarize the contributions of my work to conclude this thesis in Chapter 6.

In Part II, I present the publications and an accepted manuscript produced during the three years of my PhD that support the contributions of this thesis. They were produced under the supervision of Hans-Jörg Schulz, in which I was deeply involved:

In Chapter 7, I present a publication on the theoretical foundations of trust in VA by putting the trust continuum in the context of VA and dissecting the elements in the human-machine analytical process from the perspective of trust calibration. As the first author, I initiated the conceptualization of this work and contributed to the majority of its development and writing. This paper was presented in *2020 IEEE Workshop on TRust and EXpertise in Visual Analytics (TREX)*.

In Chapter 8, I present a publication that relates DM theories to VA and delineates the difference between making decisions "with" and "in" VA. As the sole author, I contributed to the majority of this work. This paper was presented in *2021 IEEE Workshop on TRust and EXpertise in Visual Analytics (TREX)*.

In Chapter 9, I present a manuscript on a generic guidance method that expands the conceptual space of guidance from "knowledge gaps" to "decision points" and provides a detailed step-wise guidance method through decision support. As the first author, I contributed to the majority of this work, including its conceptualization, method and prototype development, expert workshops, and writing. This paper was accepted by the *Information Visualization* journal in November 2022 with minor revision.

In Chapter 10, I present a publication on providing guidance through vibrotactile feedback with its design space, a set of prototypes, and corresponding user studies. As the first author, I contributed to the development of the design space, prototype, user studies, and the majority of the writing. This paper was presented in *International Symposium on Visual Information Communication and Interaction (VINCI) 2020*.

In Chapter 11, I present a publication on a sketchy rendering technique for aiding the recollection of information in regular visualizations with a description of this technique and its evaluation. As the second author, I facilitated the development of the prototype, designed and conducted the user studies, and contributed to the evaluation section of the writing. This paper was presented in *Eurographics Conference on Visualization (EuroVis) 2020*.

2 BACKGROUND

This thesis builds upon a range of interdisciplinary topics, including VA, DM, trust, and guidance. This chapter provides an overview of the related research in these topics to establish the background on which the contributions of this thesis rest.

2.1 VISUAL ANALYTICS, DECISION MAKING, AND TRUST

Connecting the complementary capabilities of humans and machines, VA utilizes a wide range of tools in visualization and data analysis to realize its aim of facilitating analytical reasoning. This section presents fundamental concepts in VA and related research in DM and trust that indicates critical research challenges for achieving the aim of VA.

2.1.1 *Fundamentals of Visual Analytics*

Thomas and Cook initially defined VA as *"the science of analytical reasoning facilitated by interactive visual interfaces"*, [257] which delineated the subject being studied – analytical reasoning, and the main tool being used – interactive visual interfaces. The major benefit that comes with VA, as Keim et al. [134] illustrated, is to connect the complementary capabilities of machine and human through close collaboration between them – machines can process large amounts of information and execute routine tasks efficiently, while humans can integrate their domain expertise and undertake vague and less-defined tasks – as illustrated in Figure 2.1. Keim et al. [134] also provided a more specific definition of VA: it *"combines automated analysis techniques with interactive visualizations for an effective understanding, reasoning and decision making on the basis of very large and complex data sets"*. This definition further clarified the context of VA research – analyzing large and complex data sets; its interdisciplinary nature – combining automated analysis techniques with interactive visualizations; and its ultimate aim – effective understanding, reasoning and decision making with such data.

Inspecting the definition of VA from Kiem et al. [134], the components of VA are already implied. VA is built on the basis of data sets, making *data* an essential component to start with. Meanwhile, to combine automated analysis techniques with interactive visualizations,

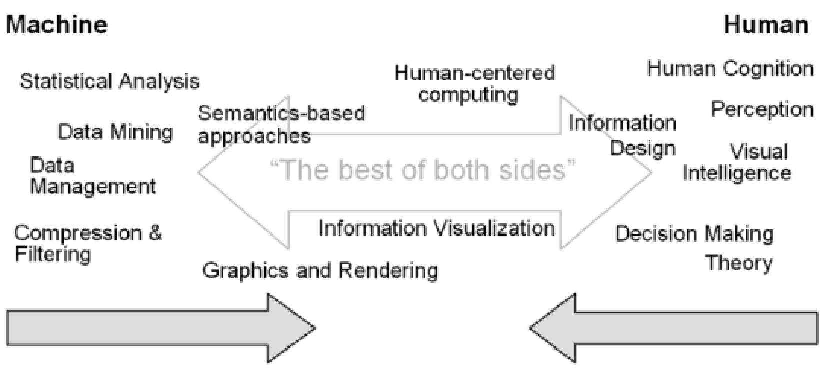


Figure 2.1: Kiem et al.'s account of the complementary capabilities from machine and human that are utilized in VA systems [134].

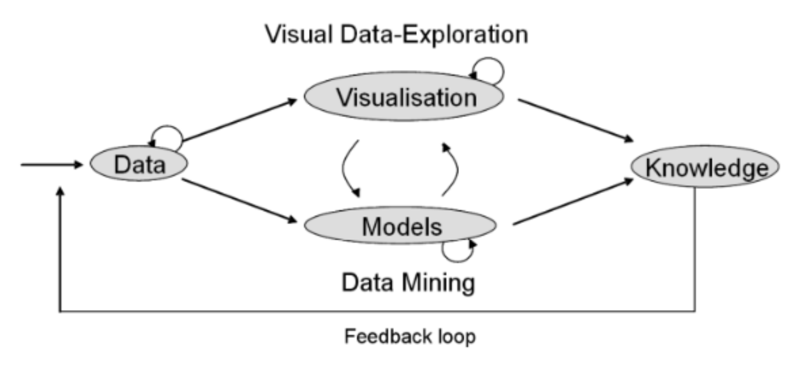


Figure 2.2: The model for VA by Kiem et al. [135], composed of Data, Visualization, Model, and Knowledge.

algorithms that enable such automated analysis and *visualizations* that present the underlying data are two important technical components of any VA system. Finally, to support effective understanding, reasoning, DM, *human user* is also a key component for VA.

Kiem et al. [135] also abstracted a model for VA that included four components – data, visualization, model, and knowledge – as shown in Figure 2.2. Another model that has been widely adopted in VA research comes from van Wijk [265], as shown in Figure 2.3, which focuses on formulating the value of visualization. His model includes some additional components – perception and exploration connected to knowledge on the human user side, specification governed by user exploration and feeding into visualization on the machine side, as well as image connecting visualization with human perception. Sacha et al. [222] further built the knowledge generation model in VA by considering the process of human sensemaking with three distinct loops – exploration, verification, and knowledge generation – as shown in Figure 2.4.

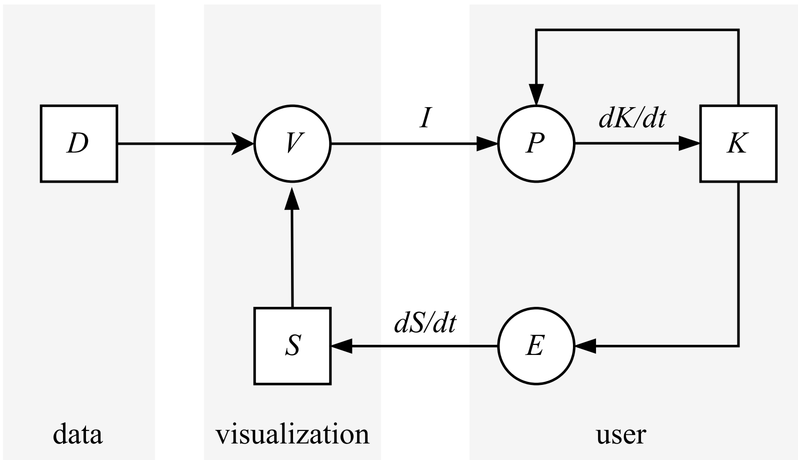


Figure 2.3: The model for the value of visualization by van Wijk [265]

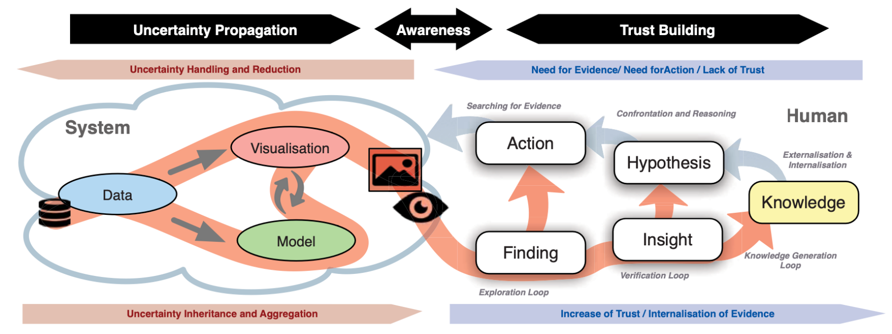


Figure 2.4: The knowledge generation model in VA by Sacha et al. [221] with the inclusion of uncertainty propagation and human trust building. The left side illustrates the components in VA systems, and the right side illustrates human reasoning process with three distinct loops – exploration, verification, and knowledge generation.

2.1.2 *Decision Making and Visual Analytics*

In the characterization of VA, Keim et al. [134] already pointed out that VA is to help people "ultimately make better decisions" and concepts in "decision-making need to be applied and extended" in VA research. Subsequent research followed suit and leveraged DM research for assisting users to better decisions with the help of VA. On a conceptual level, Padilla et al. [200] developed a cognitive framework for DM with visualizations through integrating the two types of processing in human reasoning. More specific applications and discussions have also been made. For example, FairVis by Ahn and Lin [33] aimed to promote fairer decisions with VA through identifying potential biases in Machine Learning models, and Cho et al. [48] dived into the specific bias of anchoring effect and its implications on DM with VA.

These research efforts take important steps towards supporting users to make better decisions with the help of VA. However, recent research expressed concerns about the integration between DM and VA research. Dimara and Stasko [67] underlined the lack of inclusion for DM in task taxonomies across visualization and VA research based on a review of 38 related papers. They reflected on such lack of inclusion, speculated the explanations of why these taxonomies omit decision tasks, and called for further integration of DM research in the visualization community. Their explanations were based on three possible speculations – decision making is a combination or subtask of other tasks, is too high level and not operationalizable, or too low-level to be included. Such discussions provided insights in connecting DM research with VA through Simon's Intelligence-Design-choice model [236] for DM processes, while indicating the difficulty to properly clarify the relationship between DM and VA based on existing research.

The difficulty to delineate and understand DM in VA was also reflected in recent research in data analysis. Liu et al. [157] investigated the serial nature of data analyses through empirical studies with analysts along real-life analytical processes to understand the rationales and motivations behind their decisions, which indicated the complex dynamics and implicit reasoning when human users make decisions in analytical processes. This was further exemplified by another empirical research focusing on the role of alternatives in data analysis [155], which underlined the fluidity of attention that analysts pay to the alternatives and how they often implicitly consider the alternatives and therefore the decisions.

2.1.3 *Trust and Visual Analytics*

Previous VA research related to trust has focused primarily on how to ensure the trustworthiness of machines and how human users perceive it. Building on the knowledge generation model, Sacha et al. [221] structured the trust dynamics in VA with the uncertainty propagation in VA systems, the trust building process in human users, and the users' awareness of such uncertainty in VA systems as a medium to building trust, as shown in Figure 2.4. Similarly, the review of trust in InfoVis from Mayr et al. [172] framed the two sides of trust with the trustworthiness of visualizations and the trust perception from human users. Furthermore, McNutt et al. [178] summarized and discussed how to surface the "visualization mirages" that might arise in visualization design and could potentially influence the trustworthiness of these visualizations.

These studies provide insights for designing trustworthy VA systems and ensuring trustworthy analytical results. However, the trustworthiness of machines, i.e. VA systems, is only half of the story in the human-machine analytical process – Many potential pitfalls in human sense-making and DM processes also pose challenges to the trustworthiness of human users in VA processes. These pitfalls are underlined by a wide range of research – differing perceptual abilities and various perceptual errors can distort how humans perceive information [113, 266]; lack of domain and VA knowledge of the human users leads to misinformed and untrustworty decisions based on such false or inaccurate knowledge [222]; cognitive biases give rise to conscious and unconscious prejudices that directly contribute to untrustworthy decisions [66, 74]; and situational factors such as mood and environment also create different contexts that could trigger various issues and cause unstable decision making [209, 235].

2.2 RELATED APPROACHES

The previous section presents the fundamentals of VA as well as research challenges in DM and trust for delivering VA's promise of connecting the complementary capabilities of humans and machines. The lack of clarity and understanding of DM in VA disallows machines to adapt to the serial, dynamic, and contextual nature of human decisions; while the insufficient focus on the trustworthiness of human decisions poses challenges to integrating the unique capabilities of humans with machines. Fortunately, there has been previous research in VA and DM that provides puzzle pieces promising for solving parts of the challenge. To provide further background on related research, the following presents the fundamentals of two related approaches that have shown their potentials in supporting trustworthy DM.

2.2.1 *Guidance in Visual Analytics*

Guidance in VA was initially inspired by the concept of "user guidance" in HCI. Smith and Mosier [240] in their guidelines for designing user interface software defined user guidance as "*error messages, alarms, prompts, and labels, as well as to more formal instructional material provided to help guide a user's interaction with a computer*". Such a concept has led to a wide range of research in different domains, including VA. Schulz et al. [231] made their first attempt to place the concept of guidance in the context of visualization research, which was then fully extended and initially characterized by Ceneda et al.[41] in VA to resolve knowledge gaps. They focused on the conflicts in the ever-growing complexities in data therefore VA tools and the increasing difficulties for users to achieve their analytical goals that come with such complexities.

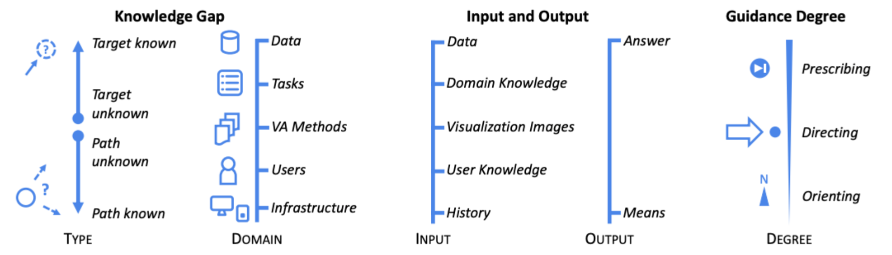


Figure 2.5: The aspects in guidance from its initial characterization in VA by Ceneda et al. [41].

Although the concept of guidance in VA was characterized around resolving knowledge gaps, the potential of using guidance to support users' decisions during the analytical process has been hinted in previous research. In the framework for designing guidance [37], the authors stated that "if the end user is not required to *take decisions* and to *reason about alternatives*, guidance is not needed." and the designers of VA systems need to "identify these *decision points* in the analysis" in order to offer guidance. This suggests that guidance could potentially help to support decisions during VA processes. To provide the background of guidance in supporting trustworthy DM in VA, the following presents its related concepts, applications, and research potentials.

2.2.1.1 *Fundamentals of Guidance*

The following presents the concepts of guidance from four perspectives – what is guidance from its characterization, why to use guidance from its goals, when to use guidance from its objectives, as well as how to use guidance from its design and practical frameworks.

WHAT IS GUIDANCE In order to resolve knowledge gaps in VA processes, guidance was initially characterized with the type and domain of such knowledge gap, input and output, as well as guidance degree [41], as shown in Figure 2.5. Based on this characterization, a conceptual model of guidance in VA was further developed based on van Wijk's [265] model for visualization to specify the relationship between guidance and the components in VA, as shown in Figure 2.6.

Knowledge gaps stem from the discrepancy between what users know and what they need to do in VA systems, which can be categorized with their *type* and *domain*. They can be generally categorized into 2 types – *target unknown* where users do not know what to achieve, and *path unknown* where users do not know how to achieve the target. From the specific domains, knowledge gaps can come from 5 domains – *data* being analyzed, *tasks* being undertaken, *VA methods* being used, as well as *users* of the VA system and the *infrastructure* it is built on.

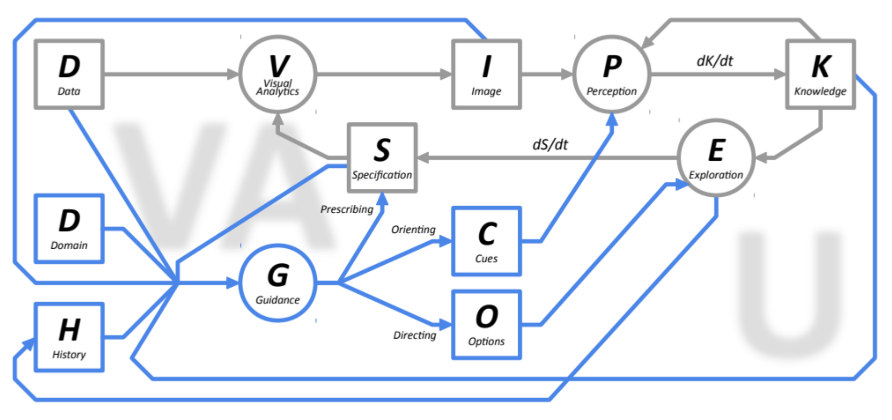


Figure 2.6: Components of guidance (in blue) from the characterization of guidance by Ceneda et al. [41] based on van Wijk's model of visualization [265] (in gray).

Input and *output* of guidance describe the basis on which guidance can be built and the way in which the guidance can be presented to users. The input of guidance can be taken from the *data* being analyzed, the *domain knowledge* stored in the system, the *visualizations* requiring different considerations and interactions, the *user knowledge* indicated by users or inferred from their actions, and *history* in the analytical paths. The output of guidance can be provided as direct or indirect *answers* and through a wide range of visual or non-visual *means*.

Finally, *guidance degrees* indicate how forceful the provided guidance is. *Orienting* guidance maintains users' mental map without explicitly indicating preferences among the possible options, such as providing suggestions, denoting changes, or presenting dynamic exploration history; *directing* guidance points users towards certain analytical directions with clear preferences indicated, which is commonly seen in recommender systems; and *prescribing* guidance pushes users through a specific analytical path, which is often computed by the VA systems or preset by designers.

WHY TO USE GUIDANCE The initial characterization [41] indicated a clear goal of guidance in VA – to *resolve knowledge gaps* that hinder users from moving the analysis forward. However, Collins et al. [51] suggested that guidance can serve a wider range of goals in addition to resolving knowledge gaps. They proposed "to inform, to mitigate bias, to reduce cognitive load, for training, for engagement, and to verify conclusions" as the goals of guidance. This extension of the goals of guidance opens up the design space of guidance to support users when they do not lack the related knowledge yet still have difficulties reaching their analytical target.

WHEN TO USE GUIDANCE In the review of guidance approaches, Ceneda et al. [39] described the analysis objectives in the provided guidance in 5 categories – *Data, Visualizations, Explorations, Models, and Verification & Knowledge Generation*. These objectives were later reframed by Pérez-Messina et al. [203] as the "when" dimension in their topology of guidance tasks to capture the contexts in which these tasks occur, including *Data Transformation, Visual Mapping, Parameter Setting, Model Visualization, Model Building, Exploration, and Knowledge Generation*.

HOW TO USE GUIDANCE To help VA designers practically apply guidance, research has been done in design a implementation frameworks for guidance. The framework from Ceneda et al.[37] proposed 4 steps – *analysis goals, knowledge gaps, guidance generation, and guidance feedback* – with considerations needed in each step for guidance design. The practical framework from Sperrle et al. [244], on the other hand, took a bottom-up approach by considering what suggestions and strategies can be used and provided a library for guidance generation.

2.2.1.2 *Related Applications of Guidance*

Despite a concept recently introduced to VA, the usage of guidance in VA has been widely adopted even before its formal characterization. Ceneda et al. provided a review of applied guidance approaches summarizing 53 related papers from the perspectives of analysis objective, guidance degrees, guidance direction, and guidance inference. [39]

Some guidance applications explicitly mentioned their goals to support DM with VA systems. For example, Migut et al. [181] provided guidance for the decision to classify psychiatric patients with relevant metrics and visualizations in the model building process. Xie et al. [284] utilized casual relations to guide decision making in what-if analysis with casual graphs.

Meanwhile, some other guidance applications focused on specific concepts in DM, such as alternative and bias. For example, Voyager from Wongsuphasawat et al. [280] provided guidance in visualizations by presenting the possible alternatives relevant to partial specifications provided by users. Boba from Liu et al. [158] offered guidance in multiverse analyses through visualizing the relevant metrics of each alternative and the sensitivity of each analytical decision. Wall et al. [191, 270] generated guidance for mitigating biases in the analytical process through detecting and making users aware of their potential biases.

2.2.1.3 *Research Potentials*

From its characterization [41], one can already observe the conceptual connections between guidance and DM. The two types of knowledge gaps, target unknown and path unknown, both address issues that occur in decisions – target unknown indicates the lack of knowledge on what to aim for, where higher-level decisions need to be made on the direction of the analysis; while path unknown surfaces the knowledge gaps in executing the analysis, where users make a series of lower-level decisions leading towards the target. Meanwhile, the different guidance degrees were also framed upon how guidance enforces its suggestions among the possible alternatives to users – orienting degree implies the alternatives are presented without preferences, directing degree communicates the preferences among the alternatives, whereas prescribing degree only indicates the "ideal" one. The extension on the goals of guidance by Collins et al. [51] also indicates its potential in supporting DM, including mitigating biases [88], easing cognitive effort [212], and providing relevant relevant information [269], which are also important topics in DM research. These connections make guidance a potential approach to connect VA and DM research for supporting trustworthy DM in VA. As indicated in the applications of guidance, a range of guidance techniques have also demonstrated its potential to support DM in VA.

Despite the potentials, guidance in VA is still conceptualized around resolving knowledge gaps, which is an important element of DM yet far from the full picture – analysts can have enough knowledge, but still be subject to various biases, limited by their cognitive ability, and disoriented in the complex analytical process, which all have a negative impact on the trustworthiness of their decision making. Therefore, the concept of guidance needs to be reexamined in the context of DM to comprehensively capture these potential issues beyond knowledge gaps and fully enable its potential to support trustworthy DM.

2.2.2 *Decision Support*

The development of Decision Support Systems (DSS) stemmed from DM and Information Systems (IS) research with a specific focus, as the name suggests, on supporting human decisions [207]. To this end, research in DSS [31] has focused on a similar set of goals that VA guidance aims to achieve, such as to provide important information, overcome cognitive limitations, and mitigate biases.

The overlap in the goals between DSS and VA guidance research makes DSS promising to be adapted in the context of VA to support trustworthy DM. Therefore, as a background for the contributions of this thesis, the following presents the fundamentals of DSS. For further

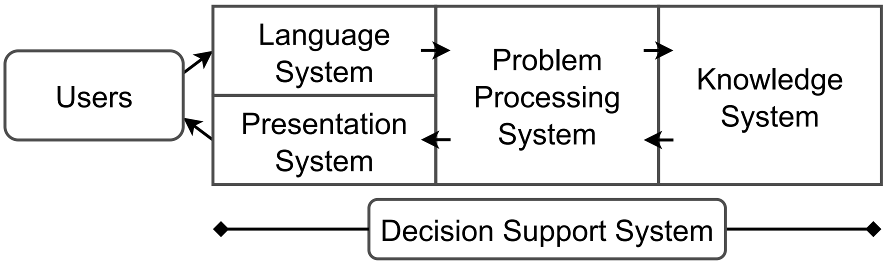


Figure 2.7: The components in the architecture of DSS based on the *Handbook on Decision Support Systems* [268].

background, the *Handbook on Decision Support Systems* [31] offers a comprehensive overview of DSS from a wide range of perspectives.

2.2.2.1 *Decision Support Systems*

Just as VA brings together the complementary capabilities of machines and humans, DSS also complement human decision makers with the powerful capabilities of IS in processing and storing large amounts of information [207]. This helps to execute decision processes that might be computationally demanding and reduce the complexity of human decisions. To lay the background for the contributions in this thesis that integrate DSS with VA, the following presents the fundamental concepts of DSS.

ARCHITECTURE A generic architecture of DSS is composed of four components – language (input), presentation (output), knowledge (database), and problem-processing (model) [268]. The representation of DSS is reflected in the first three components. The language component takes in information from decision makers, the presentation component displays information to decision makers, whereas the knowledge component stores the related information and knowledge. In addition, the problem-processing component acts as the underlying engine for DSS that integrates other components together, computationally helping decision makers to recognize and solve the decisions at hand. Figure 2.7 visually demonstrates the relationships between the components.

TYPES Progressing with the advancements in different components of the architecture, DSS can be generally categorized into five types – model-, data-, communication-, document-, and knowledge-driven [207, 268]. The earliest DSS were mostly built around a range of models in finance, optimization, and/or simulation, with the capability to structure and analyze decision problems with limited inputs. With the development of IS and data analysis, more and more DSS started to focus on storing and processing large amounts of data to support DM. Meanwhile, advancements in network and communication enabled DSS to support group DM with multiple decision makers, and increasing access to the Internet allowed decision makers to easily access a wider range of documents in DSS. Furthermore, integration of different algorithms and Artificial Intelligence allowed knowledge-driven DSS more easily suggest actions to decision makers.

BENEFITS Generally speaking, DSS have two types of typical benefits – better DM process and better outcome of decisions [206]. From a *process-oriented* perspective, DSS support decision makers throughout each stage. First, DSS facilitate decision makers to discover the decisions and help to inform them with relevant information. Subsequently, DSS allow users to better model their decisions, consider different perspectives, and easily evaluate the alternatives. Finally, DSS help to reduce decision makers' cognitive effort in the their decisions, enabling them to consider more scenarios, process more criteria, and verify their analyses – which all contribute to higher decision quality. From an *outcome-oriented* perspective, DSS can improve the quality of the decisions and bring various benefits that come with better decisions, such as reducing costs and risks, as well as improving efficiency and value.

2.2.2.2 *Research Potentials*

Research in DSS has shown great potentials to support trustworthy decisions and is closely connected to the concept of guidance in VA. From an architectural perspective, DSS [268] have similar components to VA [134]. The *knowledge* system in DSS contains various information that in VA is stored in the *data* component and informed by related *knowledge*; the *problem processing* system in DSS offers the computational components as the *models* in VA; and the user interface in DSS is composed of the *language* and *presentation* systems that often utilize similar techniques as the *interactive visualizations* in VA. These similarities make it architecturally and technically viable to integrate DSS and VA to support DM. From the perspective of benefits and goals, DSS aim to support better decisions [206] and contribute to the goals of VA guidance, such as to inform, to reduce cognitive biases & load, and to verify the analyses [51].

2.3 SUMMARY

The conceptualization of VA aimed to connect the capabilities of both humans and machines for analyzing complex data. To fully enable such close cooperation between humans and machines and unlock the power of VA, the trust between the two entities needs to be calibrated. As human decisions are consequential for the results of analysis yet subject to a range of fatal pitfalls, such decisions need to be supported for them to be trustworthy. However, previous research has focused primarily on using VA to support external decisions rather than decisions during the analytical process, and trustworthiness of machines, not human users, has been the main focus of previous VA research related to trust.

Two promising groups of related approaches for supporting trustworthy DM in VA were also presented – guidance in VA and decision support in DM. Nevertheless, neither of them comprehensively captures the challenge of trustworthy DM in VA – the concept of guidance was characterized around resolving knowledge gaps in VA, with DM slightly touched upon; while research in DSS has not been fully adapted and integrated into the contexts of VA. These challenges prompted the following contributions to connect VA and DM research from theoretical, methodological, and empirical perspectives in order to support trustworthy DM and unlock the power of VA.

3 DECISION MAKING AND TRUST IN VISUAL ANALYTICS

Assisting users to "ultimately make better decisions" [135], VA aims to enable a close collaboration between humans and machines. During such collaboration, the human users also continuously make analytical decisions based on their sense making process in order to steer the computations towards the analytical goals. For example – Which data outliers to exclude? How to process the data with different algorithms and parametrizations? What visual encoding and layout to use for visualizing the results? These analytical decisions are fundamental to unlock the power of VA, yet have a great degree of undisclosed freedom that might produce drastically different results with varying levels of trustworthiness [232].

With a wide range of common pitfalls in human decisions, the trustworthiness of these analytical decisions needs to be supported to ensure the trustworthiness of the resulting analytical insights. However, as discussed in Chapter 2, decisions in VA have often been seen as an external goal that VA systems aim to aid instead of a series of processes during the analytical sessions, and the concept of trust has been inspected from the perspective of how human users perceive the trustworthiness of machines. The lack of explicit delineation of user decisions in VA and their trustworthiness makes it difficult to properly support them.

To delineate the concept of trustworthy DM in VA, I summarize in this chapter the contributions in my two position papers that put trust and DM research in the context of VA. Further details of these two publications can be found in Chapter 7 and Chapter 8. Supplementing the related work outlined in Chapter 2, this chapter also provides the theoretical foundations for trustworthy DM in VA on which the method and techniques in the following chapters are built.

3.1 DECISION MAKING "IN" VISUAL ANALYTICS

In the position paper *Making and Trusting Decisions in Visual Analytics*, I clarified the entangled concept of DM in VA research by delineating the difference between making decisions *in* VA and *with* VA and analyzing how related theories in DM can contribute to supporting decisions *in* VA. The following summarizes such delineation and the potentials it opens up for applying the DM theories mentioned in the background Chapter 2 in VA.

I first make an explicit distinction between making decisions *with* and *in* VA – this thesis focuses on the latter. Decisions made *with* VA refers to the external decisions supported by VA systems, such as what stocks to invest in or what strategies to use for mitigating a pandemic. Decisions made *in* VA, on the other hand, refers to the decisions during the analytical sessions, such as which data areas to zoom in, how to process and analyze the data with different algorithms, and what encodings to use for the visualizations, which all implicitly lead to the external decisions eventually made *with* VA. The delineation between making decisions "with" and "in" VA not only creates new research opportunities for effectively supporting human decisions, but also enable DM theories to be further integrated with VA research.

3.1.1 *Decision Making Process*

Herbert Simon, one of the founders of DM research, proposed a widely adopted model for DM process with three fundamental stages – *intelligence, design, and choice* [236].

- *Intelligence* stage is where decision makers collect relevant information and integrate it with the context to identify the environments and conditions that call for decisions. A specific example in VA could be when users need to identify clustering groups in a data set, what processing algorithms are needed and how to use them are some of the essential information for users to make their decisions and move forward with their analyses.
- During the *design* stage, decision makers then formulate possible alternatives for the corresponding decision and evaluate these alternatives. For example, different clustering and dimension reduction algorithms in combination with possible parameters might be tested to find desirable clustering results.
- In the *choice* stage, decision makers select an alternative based on the previously generated evaluation in the specific context of the decision. Such choice in VA can be supported by presenting the alternatives and the corresponding decision criteria for the users to compare in detail.

Meanwhile, Simon also pointed out that DM is an *iterative* process – decisions makers might need to go back to the *intelligence* stage when additional knowledge is required, and to the *design* stage when none of the available alternatives is satisfactory [236].

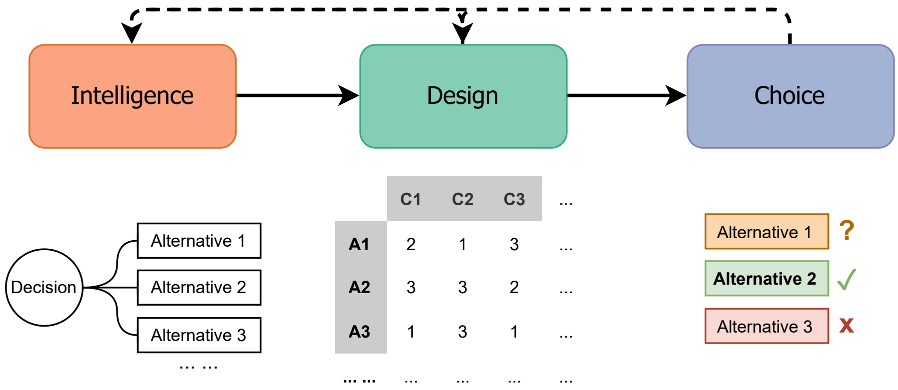


Figure 3.1: Visual illustrations of the three stages in DM process modeled by Simon [236]

Figure 3.1 illustrates this model with the three stages as well as corresponding visual explanations. Overall, this model outlines the very characteristic of decision – *a choice among alternatives*, and helps to further clarify what DM *in* VA means – when there are multiple alternatives to choose from for users to move forward with their analyses, the process leading towards the choice among these alternatives is considered DM. Applying this model of *Intelligence-Design-Choice* in VA would also help to identify, describe, and analyze human decisions in the analytical process.

3.1.2 *Complexity of Human Decisions*

Although the DM process model provides a structured manner to analyze decisions, DM is often more complex in the real world. In DM research, *"bounded rationality"* [89] is an important concept that emphasizes the limited cognitive capacity of human decision makers, which often manifests *"satisficing"* [235] behaviors – making less than optimal decisions that *satisfy and suffice* certain requirements. Furthermore, decision makers are also subject to a wide range of cognitive biases that could further distort the DM process [106].

Dual process theory of human reasoning summarizes such complexity of human decisions by considering the two types of processing of human cognitive processes [79]. Type 1 describes the automatic reasoning processes that are unconscious and intuitive, whereas type 2 refers to conscious reasoning processes based on explicit considerations. Type 1 processes are often quick and efficient but subject to unconscious biases and errors, while type 2 processes can surface such biases and errors but often take more time and cognitive effort [79]. An example of these two types of process in data visualization was given by Padilla et al. [200] presented in Figure 3.2 – when users identify if the average of two bars is closer to 2 or 2.2,

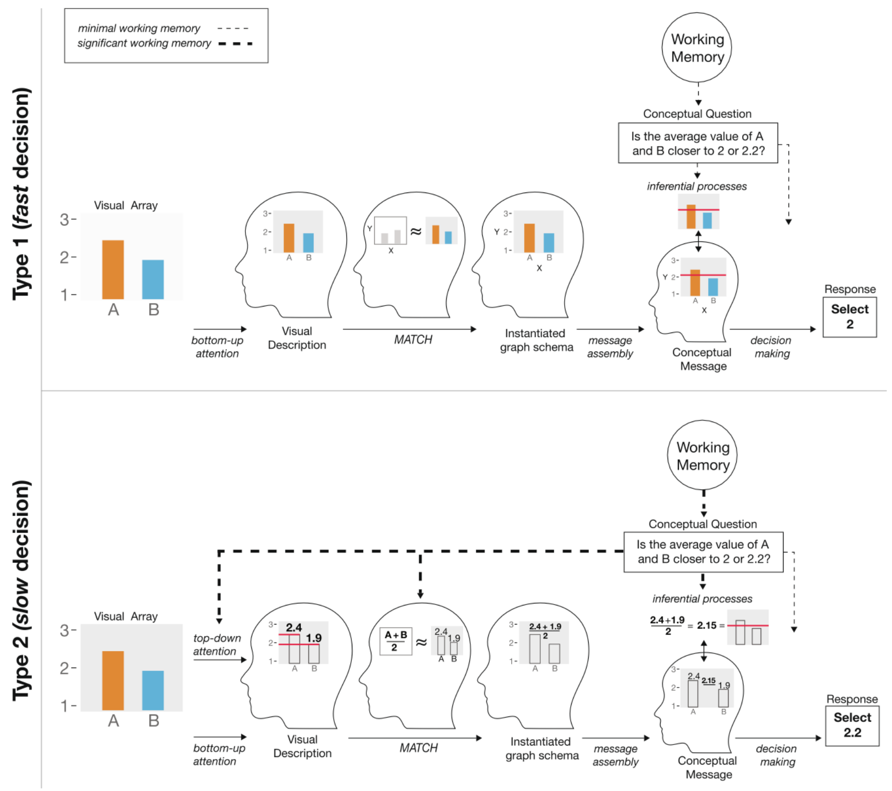


Figure 3.2: An example of type 1 (top) and type 2 (bottom) reasoning in dual process in visualization by Padilla et al. [200]. In this example, the task is to decide if the average of the bars A and B in the visualization is closer to 2 or 2.2.

they might visually estimate (type 1) or consciously calculate (type 2) the average.

For making decisions in VA processes, an example of the two types of process could be where users decide which algorithm to use based on some characteristics of the data at hand – they can base their decisions on the visual perception of data patterns (type 1) or statistical calculations of these characteristics. Both types of process are useful in different manners – type 1 processes allow quick experimentation and incorporate implicit human expertise that can be hard to articulate, whereas type 2 processes give rise to evidence-based decisions with grounded reasoning and conscious trade-off. Therefore, supporting both types of reasoning process is important in order to ensure trustworthy DM in VA.

3.1.3 *Decision Strategies*

During the *design* stage of DM, alternatives are developed to ensure the process is informed with a variety of possibilities and perspectives. Meanwhile, with an increasing number of alternatives, decision makers can be overwhelmed with too many options to evaluate in the *choice* stage. Different strategies, generally divided into compensatory and non-compensatory ones, are developed to evaluate alternatives and make the choice [61, 76, 151, 192].

Compensatory strategies consider the trade-offs between the conflicting attributes of the alternatives, where all available attributes are integrated together to evaluate each alternative [192]. For example, additive strategies are commonly used to combine the attributes through a weighted sum [61]. These strategies are often applied in less constrained contexts with sufficient information, time, and resources, while providing the benefit of counteracting potential risks and cognitive biases with their more structured and comprehensive evaluation processes.

Non-compensatory strategies, on the other hand, do not consider these trade-offs, but inspect the attributes individually, often with certain heuristics and rules [192]. For example, elimination-by-attributes is a classic non-compensatory strategy that eliminates the alternatives that do not meet certain requirements in the corresponding attributes [151]. These strategies are usually used in more constrained contexts where available information, time, and/or resources are limited.

In practice, including most decisions in VA, both groups of strategies are often used together. For example, heuristics can help users narrow down which processing algorithms are appropriate for the data at hand, while further evaluation of these algorithms in their performance can help them to decide which one to use. Research in DM indicates that even simple non-compensatory heuristics can lead to highly accurate decisions [170]. Therefore, VA systems can support trustworthy DM by incorporating these strategies in the design and prompting users to make their decisions based on explicit criteria and heuristics.

An important similarity between the two types of strategies is that they both process the alternatives based on a range of attributes, i.e. criteria [76]. Such similarity allowed a group of approaches – *Multi-Criteria Decision Analysis* (MCDA) [237] – to integrate both types of strategies through considering all or parts of the available criteria. MCDA models combine a range of criteria based on their corresponding values/utilities, outranking relationships, and/or decision rules to evaluate the alternatives [224]. Specifically, functional approaches combine the criteria together through weights and value/utility functions [279]; outranking approaches choose, rank, or sort the alternatives based on the comparisons between them in each criteria [22]; and decision rules set specific logics and conditions to eliminate and evaluate the alternatives [182].

3.2 TRUSTWORTHY DECISION MAKING IN VISUAL ANALYTICS

With the focus of "calibrating" trust between humans and machines along a trust continuum instead of "building" towards full trust, the position paper *Beyond Trust Building – Calibrating Trust in Visual Analytics* analyzes the potential trust issues in both machines and human users. In the following, I summarize the contributions of this paper on the trustworthiness of human decisions in the context of VA.

Defined as *"the belief that the trustee will act in the best interest of the truster in a given situation."* [169], the concept of trust captures the mutual dynamics between these two connected entities. When we think about the relationships between humans and machines with VA systems, the human users are often the ones who have certain interests at stake – to invest money, make plans, and diagnose patients, while machines are the ones that facilitate users and protect their interests. Therefore, it does not come as a surprise that trust has been mostly understood as a one-way street of how human users trust machines in previous VA research [59, 172, 221]. However, as previously discussed, human users are subject to a range of pitfalls that can severely undermine the trustworthiness of analytical results. Meanwhile, with the rise of mixed-initiative VA [166], human users and machines are increasingly seen as peers with a mutual relationship. In the following, I reflect on the potential issues and summarize the existing work related to the trustworthiness of human decisions in VA through four components – perception, knowledge, judgment, and context (situational state). Figure 3.3 shows the overall structure of components that influence trust dynamics between human users and VA systems included in this paper.

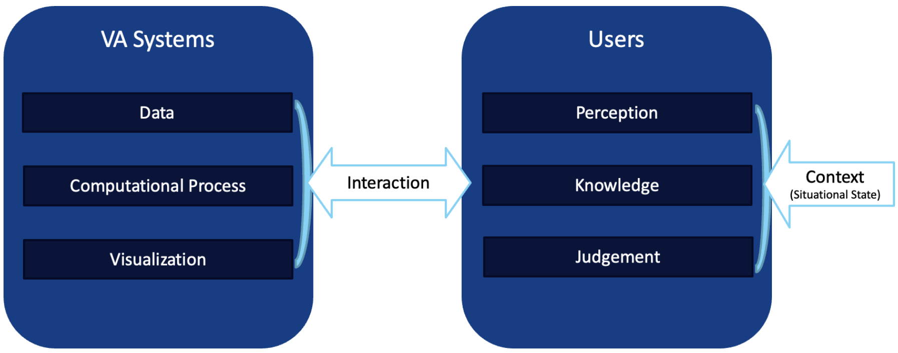


Figure 3.3: Components influencing the trustworthiness of human users and VA systems.

3.2.1 *Perception*

Perception is an important component of human decisions VA, as it is the vital connection between the output of VA systems and the input to human reasoning [221]. The perception of the information presented in VA systems also builds the foundation on which the human users make their decisions. Therefore, ensuring trustworthy human perception is essential as to support trustworthy human decisions in VA.

Early visualization research probed into perceptual errors that directly distort how human translate visual information, such as line width and sine illusions [113, 266]. Particularly in VA, the interactive nature of the user interface can make it difficult for analysts to be aware of all important changes, such as animations and updates in the visualizations. Such phenomena of humans' failure to perceive visual changes are termed as "change blindness". Perceptual abilities can also limit the trustworthiness of the information extracted from visualizations. For example, color is a commonly used encoding in data visualizations, while human sensitivity to color often decreases with age and colorblindness [234].

These challenges in human perception have been addressed mainly through visual adaptation and emphasis. For example, "change blindness" can be mitigated through underlining the changes and drawing users' attention, such as morphing, crossfading, and wireframes [195]. Furthermore, utilizing developed tools [105] in color selection helps to ensure analysts with different perceptual abilities can successfully differentiate between colors and correctly perceive the intended information.

3.2.2 *Knowledge*

When conducting analyses in VA systems, users understand and internalize the information they perceive based on their existing knowledge construct. Misinformed and/or insufficient knowledge will lead to not only misunderstanding of the perceived information but also misuse of the VA systems.

The lack of expertise with VA methods and techniques, such as how different data samplings and transformations works, severely undermine the trustworthiness of relevant decisions. Furthermore, just as in any digital system that involves interactions between human users and computers, users also familiarize themselves with and learn about how to use the specific VA system at hand over time. Such knowledge is also important for users to know what options are available, where to find them, and how they should be enabled. In addition to VA knowledge, users' domain knowledge of the specific dataset being analyzed is also essential for them to make trustworthy decisions. For example, the domain knowledge about how different physical measurements influence patients' health is essential for analysts to select the relevant data dimensions for their analyses.

To address the challenges in users' lack of knowledge, an evident solution is to provide users with the knowledge relevant to their decisions. Knowledge-assisted visualization helps to address the lack of knowledge by incorporating and storing the relevant knowledge in the visualizations and VA systems [123]. Following the same rationale, VA guidance also bridges the knowledge gaps that users encounter when undertaking their analytical tasks by providing relevant knowledge [41].

3.2.3 *Judgment*

The concepts of judgment and DM are deeply interconnected. Judgment appertains to the thought, opinion, or evaluation of given stimuli [30], which is closely connected to humans' cognition from within and underlines the subjective and internal cognitive processes during DM.

As discussed in Chapter 2, a range of cognitive biases in human judgment [66] negatively influence the trustworthiness of human decisions in VA. Some examples of specific cognitive biases have been discussed in existing VA research. Confirmation bias, which describes the tendency of human judgment to focus on information consistent with their existing beliefs [190], can cause users to choose datasets and algorithms that confirm their assumptions; Anchoring bias, which uncovers the inability of people to adjust their initial response even when contradictory evidences arise [48], can make users prone to maintaining their early decisions and default settings; Selection bias, which refers to situations where the sample selected for analysis is not representative of the whole dateset [96], would also severely undermine the trustworthiness of the produced results.

Existing techniques to address cognitive biases in VA have mainly focused on detecting and reminding users of these biases, such as showing the representativeness of their selections [96, 171], detecting user biases through systematic metrics [270], and uncovering the tendency to explore certain areas of data [191]. Wall et al. also outlined a design space for mitigating bias in VA systems with existing and emerging approaches [271].

3.2.4 *Context*

The same VA system can also be used in different contexts – Users might work on high performance cloud services that can deliver complex computational results in seconds or a personal computer that might take hours to complete a simple data transformation; They also might be situated in environments with different types of device, sizes of screen, and levels of noise; The analysis might also come with different goals, such as exploring possible explanations, confirming assumed hypothesis, and presenting analytical results [229]. In HCI research, context of use is an important topic that focuses on the actual conditions under which the designed system is used [164], such as the available resources, the physical environment they are in, as well as the tasks and goals they aim to achieve [248]. In InfoVis, situated visualization is also an emerging concept that takes into account the context in which visualizations are used [24].

Various contextual factors can influence how users make decisions in VA. Internally, the situational state of the users is directly connected to their DM. For example, people are more likely to make unnecessary changes in a negative mood, or settle with a less-than-ideal option in a positive mood [209, 226]. Furthermore, feeling of fatigue and threat can also lead to an increase in decision error and over-conservativeness, respectively [210]. Factors in the social and physical environment are also influential to one's DM. For example, shared offices and distractions can make it harder for users to focus on their analyses, leading to an increase in perceptual errors and inattentiveness to their biases.

Many contextual factors can be tracked and/or inferred through various inputs, such as facial expressions, eye gaze, and electroencephalogram to infer users' mood and intention [238]. Contextual information can also be directly inquired by asking users about the contexts they are in before or during the analysis. The tracked or inquired information can then be used to support user decisions adaptively.

3.3 DISCUSSION & REFLECTION

Conceptually connecting research in VA with DM and trust, the contributions in this chapter lay the foundation for the following work in this thesis. Regarding DM, an explicit delineation of making decisions *in* VA helps to put a clear focus on the user decisions made during the analytical sessions, while the inspection of relevant DM theories provides inspirations on how to support these decisions. As for the trustworthiness of human decisions, laying out the important issues and existing VA research in each component of human DM not only clarifies what should be supported in order to ensure trustworthy DM in VA, but also establishes the foundations to build methods and techniques for providing such support. Meanwhile, as many concepts and perspectives in this chapter are still relatively new, further research is needed to fully integrate them into the research landscape of VA .

The contributions discussed in this chapter also capture some important research trends. In particular, the emerging concept of mixed-initiative VA takes humans and machines as peers on par with each other [166], emphasizing the possibilities for machines to take initiatives and drive the analytical process. To enable mixed-initiative VA, machines also need to understand the actions of users and infer their intentions [246]. This makes it increasingly important for machines to be able to evaluate the trustworthiness of human decisions. Collaborative VA is another relevant research area in which the trustworthiness of human decisions plays an important role. As multiple analysts are involved in collaborative VA [108], it is not only important for analysts to understand and trust decisions from each other, ensuring trustworthy decisions from all involved analysts also becomes even more challenging. Furthermore, a range of research in VA [11, 197, 198, 223], automation [47] and HCI [78], also resonated with the discussions in this chapter on the trust calibration and trustworthiness of human decisions.

4 GENERIC GUIDANCE METHOD TO SUPPORT TRUSTWORTHY DECISION MAKING

In Chapter 3, I have inspected the components in the trustworthiness of human decisions and related DM theories for supporting it in VA. This helps to understand which human decisions and components in their trustworthiness need be supported, and provides theories on how to analyze and potentially support these decisions. To support trustworthy decisions in a generic manner, however, a practical method that connects these theoretical foundations with applications is still needed.

In recent years, the concept of guidance has been established to support users' analytical tasks in VA research, especially when users need to *"take decisions"* and *"reason about alternatives"* [41]. This makes VA guidance a potential basis for building a generic method to support trustworthy DM. However, VA guidance has been conceptualized around *"resolving knowledge gap"* [41], which admittedly is one important element in the trustworthiness of human decisions, nevertheless only provides a partial view of it. Meanwhile, research in decision support systems connects the theories and applications of DM [31], making it a potential source of inspiration for my research goal in this thesis.

Aiming to develop a generic method for supporting trustworthy DM in VA, this chapter presents a generic guidance method through decision support, which dissects user tasks in VA as DM processes and analyzes users decisions with MCDA models. In the following, I summarize the reframing of guidance as supporting decision points, describe the produced method with its development process, and reflect on the contributions. Further details of the method can be found in the accepted manuscript in Chapter 9.

4.1 GUIDANCE IN THE DECISION MAKING PROCESS

To develop guidance through supporting DM, the users tasks that guidance aims to facilitate need to be dissected as DM processes. The reflection on decision tasks in visualizations from Dimara and Stasko [67] analyzed existing task taxonomies with the *Intelligence-Design-Choice* model for analyzing DM processes from Simon [236]. To closely relate my work with existing research, I also dissected

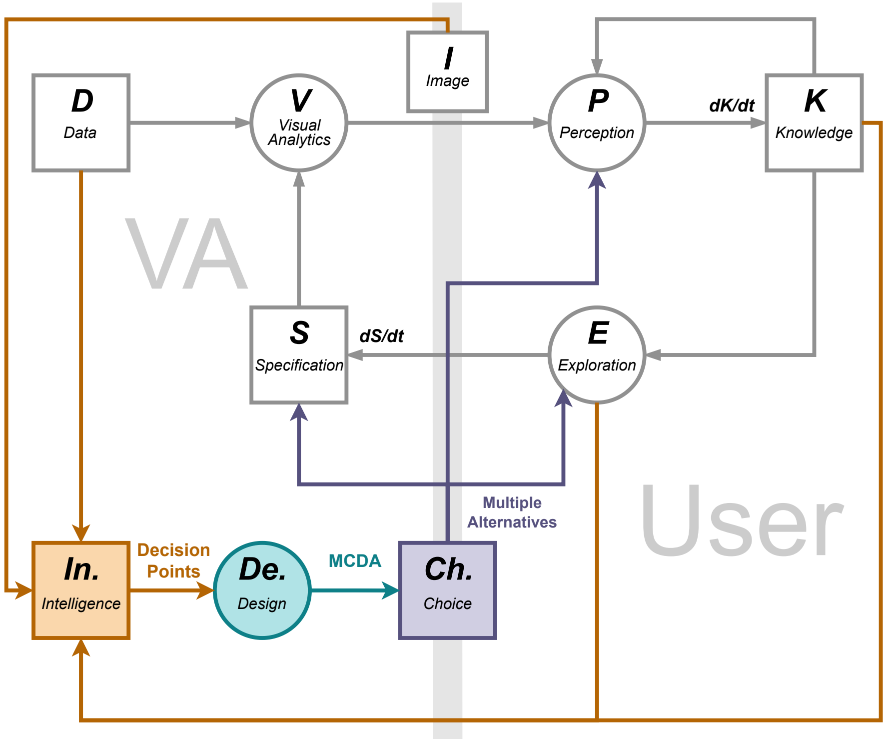


Figure 4.1: The conceptual structure of the guidance method in this chapter based on the guidance model by Ceneda et al. [41].

guidance in VA as decision support with this model. An integrated model of VA guidance with DM process is illustrated in Figure 4.1.

The first stage of the model is *intelligence*, where the environments calling for decisions are identified. In this stage, information about the decisions, such as existing knowledge, goals, and other contextual factors, is collected to detect the need for decisions and inform the later stages of DM. Existing VA research has also conceptualized guidance with a similar starting point. The initial characterization of guidance in VA frames a range of components, such as data, user knowledge, and visualization images, as the input for guidance to identify and support the knowledge gaps [41]. The framework for guidance from Ceneda et al. [37] also initiates the process of designing guidance through identifying users' analysis goals and knowledge gaps. The *intelligence* stage and the input of guidance are clearly aligned in the aim to identify scenarios where an action (to decide from the users' perspective and to guide from the systems' perspective) is needed and to collect information to support the action.

The second stage, *design*, is concerned with the development and evaluation of alternatives to later choose from. In this stage, the previously collected information is combined and analyzed for developing possible candidates to consider, and these candidates are evaluated to further inform the decision. In VA research, this is conceptualized as guidance generation, where the input of guidance is computed to generate suggestions to users – Ceneda et al. [37] outlined this as "algorithms and procedures to calculate guidance" in the "guidance generation" step of their guidance framework, and Collins et al. called it the "compute" building block for implementing guidance [37].

In the third and final stage, *choice*, decision makers then inspect and compare the developed alternatives in order to finalize their decisions. For VA guidance, this relates to the output of guidance where it is presented to users through signifying some form of evaluation of the alternatives. Particularly, in the decision tree for choosing appropriate guidance degree, Ceneda et al. [38] discussed how suggestions from guidance are made according to different guidance degrees – Prescribing guidance is presented with only one option, essentially the "best" alternative from the evaluation; directing guidance is presented with multiple prioritized options, highlighting the ranking produced from the evaluation of the alternatives; while orienting guidance simply presents the options or state, providing the information regarding the alternatives to compare and choose from.

Although these three stages are structured in sequential order, the process of DM is often iterative and dynamic – decision makers often go back to the *intelligence* stage if more information about the decisions is needed to develop and compare the alternatives, or to the *design* stage if none of the developed alternatives is satisfactory. This is also common in the context of VA – users often need to take a step back and collect more information about the dataset when they are unsure about what the data dimensions represent, or to further explore and evaluate additional algorithms when none of the available ones produces ideal results. Such dynamic nature of these analytical decisions requires the provided guidance to support them adaptively.

4.2 GUIDANCE METHOD THROUGH DECISION SUPPORT

This section provides an overview of the guidance method along with their development process. To provide context for how the guidance method was developed, I describe the process in retrospect from its conceptualization to finalization. Thereafter, I summarize the guidance method with the three stages of the DM process – *Intelligence-Design-Choice* – as well as the steps and components in each stage.

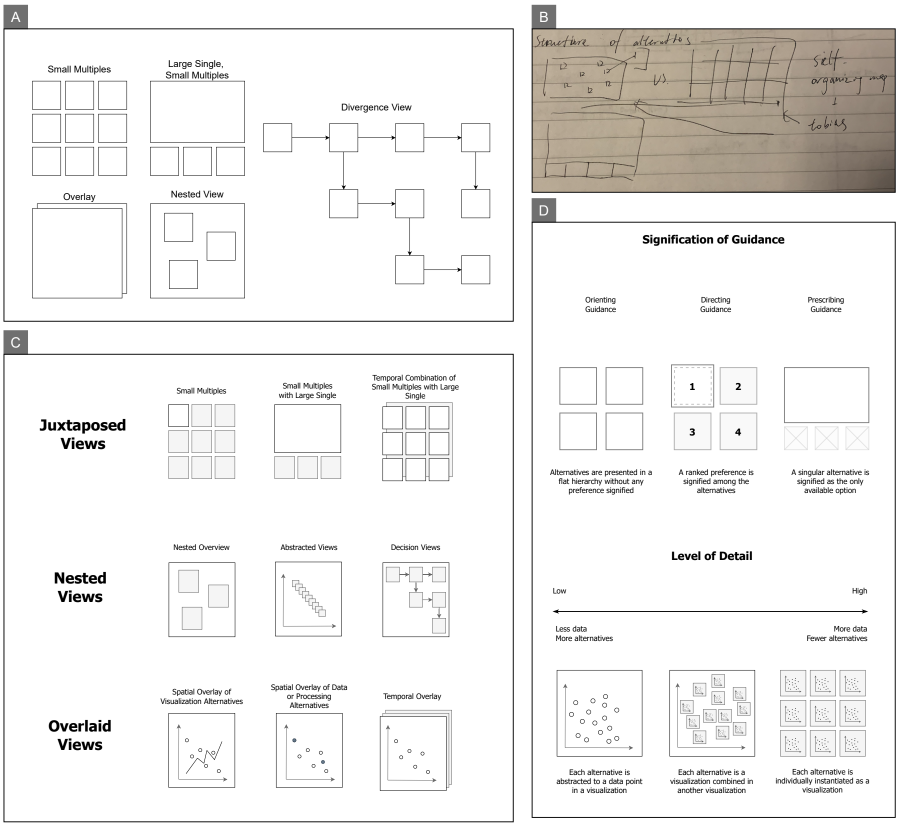
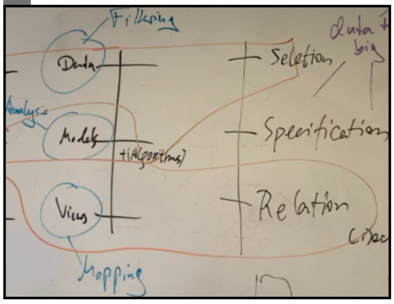


Figure 4.2: The drafts for presenting multiple alternatives in decision points: **A.** Early summary of the design patterns in existing VA research for presenting multiple alternatives. **B.** Sketches of the relationships between the different design patterns. **C.** Intermediate iteration of the design patterns based on research in multiple views. **D.** The final version for presenting multiple alternatives adapted to important concepts in VA guidance.

4.2.1 *Development Process*

The concept of trust entered the scope of my research from a fairly early stage. The initial conceptualization of this work came from the idea of exposing users to multiple alternatives in order to support their analytical decisions, as considering possible alternatives is cited in related research as an effective way to reduce biases and improve decision quality [111, 192]. Therefore, I first dived into the presentation of multiple alternatives with the aim of making users aware of possible alternatives and supporting users to compare them. During the ideation, I reviewed related work in multiple views, including general frameworks and surveys [44, 126, 213] as well as specific applications [46, 167, 272, 277]. Thereafter, I summarized the design patterns for presenting views of multiple alternatives, which informed the subsequent development on the presentation of alternatives in the *choice* stage of the produced guidance method.

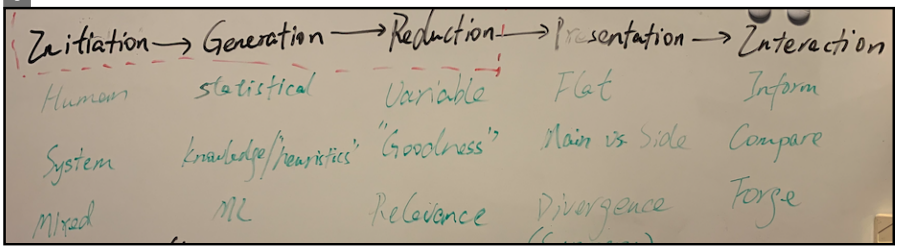
A



B

Types	Selection	Specification	Relation
Data	Navigation	Organization	Relation
Model	Algorithm	Parameter	Combination
Visualization	Chart type	Encoding	Layout

C



D

	Data	Algorithm	Visualization	Reasoning
Which	Which dataset or which subset of the data to use?	Which algorithm to use to process the data?	Which primary visual encoding to use?	Towards which end should the analysis aim?
How	How to clean and transform the data before they are used?	How to parameterize the algorithms?	How to spec the visualization?	How to compose other decision points together to reach the aim?

Figure 4.3: The drafts for the step in the guidance method to identify decision points: **A.** Early ideation of possible elements to include based on domains and types of VA tasks [229]. **B.** A refined version of the early ideation. **C.** Early draft of the overall process for supporting decision points in VA. **D.** The final version of the approach to identify decision points adapted from the interaction taxonomy by Landesberger et al. [147]

Figure 4.2 presents these iterations of the design patterns in the guidance method for presenting multiple decision alternatives.

After realizing the connection between DM and the concept of VA guidance, I proceeded to consider how to structurally and comprehensively detect the decision points where guidance is needed. As the concept of decision points is new to research in VA guidance, I developed my ideas through inspecting general VA and DM research, such as task taxonomies [23, 67, 97, 109, 147, 229] and decision analysis [13, 111, 192], and finally settled on a structure adapted from the taxonomy of VA interactions by von Landesberger et al. [147]. Figure 4.3 illustrates selected drafts produced during the development of the method to comprehensively identify the decision points in VA. This particular taxonomy was chosen as the basis for detecting decision points for three main reasons: First, the universality of this taxonomy has been validated in their paper through comparing it with a number of existing taxonomies. Furthermore, its high level of abstraction makes it easier to focus on decisions that might significantly alter users' analytical paths. Finally, the inclusion of reasoning tasks uncovers possible interactions between the series of decisions in analytical processes that can influence each other. These considerations led to the development of the core in the *intelligence* stage of the guidance method for identifying the decision points.

To provide a generic mechanism for generating guidance as decision support, I dived into the literature on decision analysis and DSS. Meanwhile, my previous research in DM summarized in Chapter 3 also provided me with background knowledge on this topic. After comparing different techniques and methods, I decided to use MCDA models [16, 224, 237] as the underlying mechanism for guidance generation for its flexibility to take in any quantitative metrics and its technical availability across many programming languages. Then I integrated MCDA models with VA guidance through guidance degrees, which later became the foundation of guidance generation in the *design* stage of the method.

Inspired by research on visualization design worksheets [174, 175], I developed a set of design worksheets for the guidance method to help VA practitioners apply it. These worksheets were developed through internal discussions and refined with three expert workshops. The workshops were conducted individually with each expert, where they were first introduced to the method, then walked through our method with a use case of their choice from their previous work, and finally drafted a pen-and-paper prototype for the guidance design. During the workshops, I had the opportunity to obtain valuable feedback that helped to refine both the details of the method and the presentation of the worksheets. The detailed guidance method, final set of worksheets, and accompanying prototype can be found in Chapter 9.

4.2.2 *Overview of the Guidance Method*

Connecting the *Intelligence-Design-Choice* model with VA guidance, I developed the guidance method to support trustworthy DM accordingly. The structure of these stages, steps in each stage, and main elements in each step is illustrated in Figure 4.4.

In the *intelligence* stage, *context of use analysis* from HCI research, *task & interaction taxonomy* from VA research, and *risk analysis* from project management are adapted to recognize the decision points in VA and assess their need for guidance support. The method starts with analyzing the context of use to provide the information of the environments of these decisions. With the context analyzed, VA de-

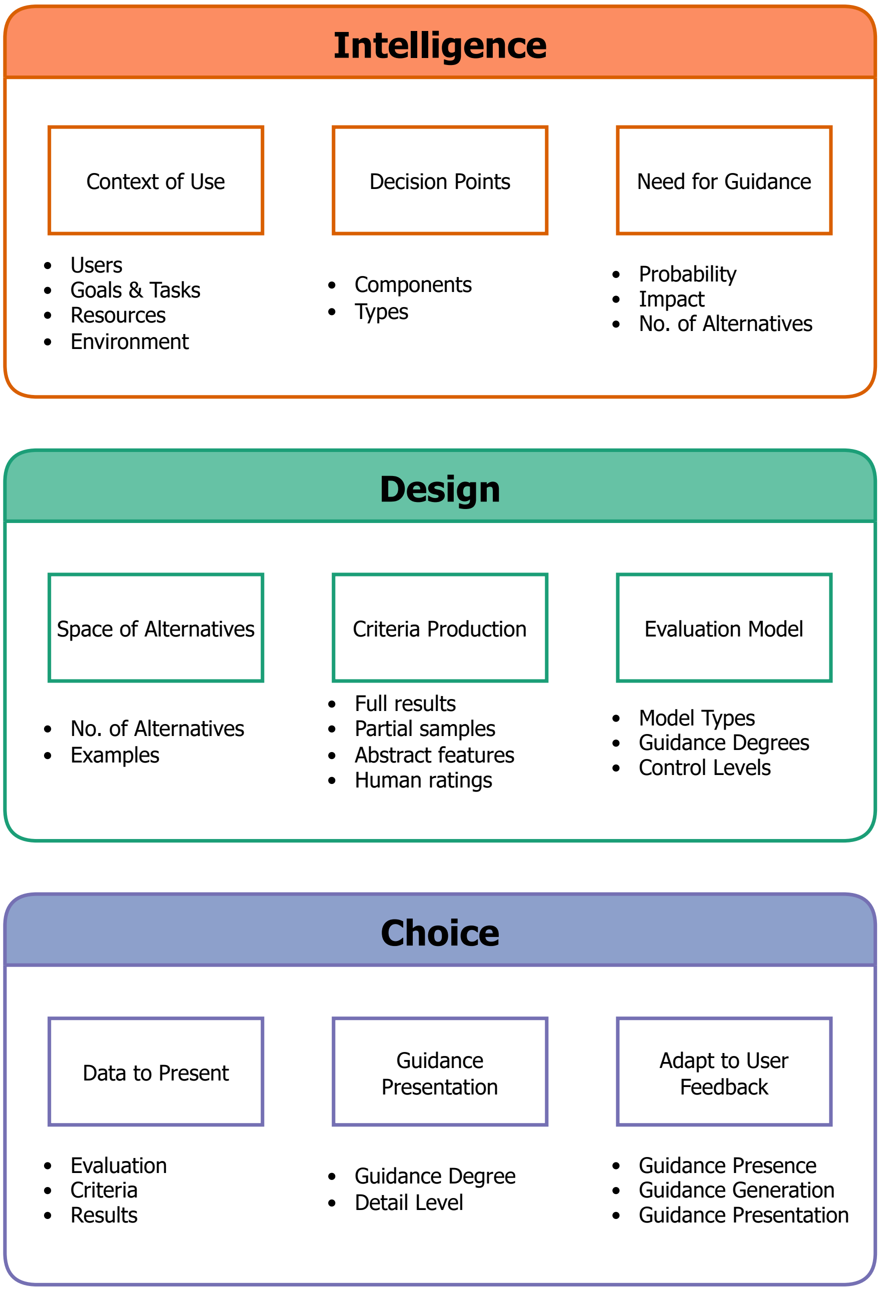


Figure 4.4: An overview of the guidance method with its three stages and the sub-steps in each of the stage.

signers then identify possible decision points where alternatives exist across the components of VA and types of decisions. These decision points are then analyzed and prioritized in their need for support through a scheme inspired by risk analysis in project management.

The prioritized decision points are then fed into the *design* stage, where the alternatives in each decision are developed and evaluated with *MCDA models* to generate guidance. The space of alternatives is first recognized for each decision point through the number and example of alternatives. Then criteria for evaluating these alternatives are produced through relevant metrics with different bases, including full results, partial samples, abstract features, and human ratings, according to the context of the decision. Three types of MCDA models – functional, outranking, and rule-based – provide the underlying mechanism for guidance generation through combining the produced criteria and evaluating the alternatives. The final step in this stage outlines how to generate different degrees of guidance and levels of user control with these models.

Thereafter, the *choice* stage takes in the generated evaluation and communicates the provided guidance with *views of multiple alternatives*. First, the generated evaluation of these alternatives, the criteria that produced the evaluation, and potential results that might have been computed to elicit the criteria are recognized as the basis of guidance presentation. Thereafter, the alternatives are visually composed to present the guidance with the recognized data according to the guidance degree and the level of detail that the alternatives should be presented in. Finally, to ensure the provided guidance is adaptive to the dynamic nature of user decisions, considerations are made for adapting the presence, generation, and presentation of guidance according to user feedback.

4.3 DISCUSSION & REFLECTION

Conceptually connecting research in VA and DM, the generic method in this chapter reframes guidance as decision support and provides practical steps to follow for designing guidance that supports trustworthy DM in VA. Such reframing enables VA guidance to utilize DM process and MCDA models. The produced guidance method also makes users aware of the decision points and formally analyzes these decisions with specific criteria and evaluation.

4.3.1 *Connecting VA and DM Research*

Reflecting on the theoretical foundations in Chapter 3, the guidance method in this chapter covers various theories in DM and components in trust.

For the theories in DM, a DM process model for analyzing decisions and MCDA models that allow different decision strategies are integrated in the guidance method. To apply the DM process model in the method, I followed the delineation of DM *in* VA and focused on dissecting the analytical decisions during VA sessions. This provides the first structural connection between DM processes and VA tasks, helping to clarify what decision tasks encompass in VA. Meanwhile, MCDA models also enable decision support with both compensatory and non-compensatory strategies – Functional approaches allow compensatory strategies that aim to combine all available information and synthesize the trade-offs between different criteria; while outranking approaches and decision rules both can be used when only partial information is considered with partial outranking relations or rules set for only a subset of criteria. This allows the produced method to practically utilize existing MCDA implementations [16, 141, 224] and flexibly adapt to various use cases that require different decision strategies.

For the components in the trustworthiness of human decisions, a wide range of trust issues was also covered in this method. Expanding on the initial characterization of guidance with resolving knowledge gaps, this method naturally includes potential issues in users' lack of or inaccurate *knowledge*. The DM process and MCDA models also facilitate users' *judgment* process of synthesizing the knowledge, which supports user decisions with specific criteria. Furthermore, through the context of use analysis, the potential issues caused by the *context* are also systematically captured with a standardized approach.

4.3.2 *Reflections on the Concept of Guidance*

Looking back on the process of developing this work, I went through several iterations of pivoting my research in different directions – from the initial idea of exposing users to diverse alternatives, then the focus on the detection of the need for guidance, to the final reframing of guidance as decision support. These iterations not only helped me to leverage interdisciplinary insights from various research fields towards the aim of supporting trustworthy DM, but also provided unique perspectives and practical tools that enriched the concept of guidance in VA.

The reframing of guidance expanded the goals of guidance from resolving knowledge gaps in its initial characterization to supporting decision points, enabling the concept of guidance to effectively cover a range of scenarios where users are well-informed and familiar with VA systems, but still would greatly benefit from guidance. For example, to be reminded of important information, to keep track of and verify their analyses, as well as to be made aware of their potential errors and biases.

Furthermore, fusing insights from existing studies in multiple views and task taxonomies with the guidance method, I also bring out new connections between guidance and these existing VA research. This helps to enrich the concept of VA guidance and allows research in VA guidance to more easily utilize these existing works in VA.

Finally, with the design worksheets, expert workshops, and an exemplary prototype, I also obtained some initial insights on the practicality of applying this guidance method. Meanwhile, more empirical studies with designers and end-users of VA systems can help to further validate if and how these tools proven in DM research support trustworthy human decisions through VA guidance.

4.3.3 *Limitations and Challenges*

With the contributions above, it is also worth noting that formalizing guidance to support analytical decisions can not solve all issues in the trustworthiness of human decisions.

First, although the structured tools in this method have been cited in DM research to support DM and improve the quality of decisions [31, 192], the criteria and DM models behind the provided guidance are still implemented by VA designers. The potential lack of knowledge and biases from these designers could feed into the provided guidance, systematically guiding VA users to make ill-informed decisions. Therefore, these criteria and models need to be carefully designed and well-researched to ensure that the guidance they provide is trustworthy.

Furthermore, for some more heuristic-driven or exploratory decisions in VA, it might be challenging to support them with specific criteria and DM models. For more heuristic-driven decisions, such as proven workflows for certain analyses or parametrizations for certain algorithms, guidance can be directly provided to users based on these heuristics without formally modeling the decisions. For more exploratory decisions, such as exploring different data areas and processing results without a specific goal, the provided guidance might be more effective if it focuses on helping users to keep track of the explored alternatives and reduce cognitive load rather than structurally analyzing specific decisions.

Finally, the awareness and structure provided by the guidance method still can not prevent users from consciously and strategically manipulating the DM process and externalizing their biases. Existing guidance approaches often orient users and direct them towards certain alternatives, where users can still refuse to follow the provided guidance and enact on their potentially biased decisions. Meanwhile, prescribing guidance that directly enforces one option for users' decisions is rarely used in practice [39], as it does not provide flexibility to adapt to users' different goals and deprives users of their sense of agency. This still stands as a societal challenge that goes beyond the design of VA systems, while existing research in social psychology cites exposure to diversity, practice of empathy, and other anti-bias training as important tools for reducing personal biases [138, 271].

5 ALTERNATIVE TECHNIQUES FOR TRUSTWORTHY DECISION MAKING

With the word *"visual"* in its name, research in VA has primarily focused on using *visual*izations, *visual* interfaces, and *visual* cues to facilitate the process of data analysis. Naturally, with the growing complexity of data and analytical process, visual complexity of VA systems also greatly increases. In these complex VA systems, it can be challenging to find available visual space and encoding to provide additional decision support to users. The visual complexity also adds to users' cognitive load during their analyses, which negatively influences the quality, including trustworthiness, of user decisions [62, 161, 188]. Meanwhile, to provide formal decision support, adding even more visual elements with complex structures might not only exacerbate the issue of increased cognitive burden, but also lead users to confuse the original VA elements with added support or even overlook some vital information relevant to their decisions [195]. This would be counterproductive to the initial aim of supporting user decisions.

To address these challenges, alternative channels of communication could be of help – an alternative perceptual channel for decision support can create clear separation with existing VA elements and emphasize its importance, while utilizing existing elements in VA systems in an alternative manner as decision support can also help cut down visual clutter and cognitive load. To this end, this chapter includes two novel techniques through alternative channels of communication that could be used to support trustworthy decision making in VA – vibrotactile guidance that utilizes an alternative perceptual channel, and sketchy rendering that utilizes existing visualization grids in an alternative manner. For each technique, the following sections outline their design space, describe the processes and results of corresponding user study, and discuss the findings with regard to supporting trustworthy decision making in VA. Further details of these two techniques can be found in the two publications in Chapter 10 and Chapter 11.

5.1 VIBROTACTILE GUIDANCE

Although VA guidance has not been limited to the visual channel in its definition [41], previous research has not touched on how to utilize modalities other than the visual one. However, as VA systems are often already visually loaded to communicate complex data and analytical information, it can be difficult to find available visual means to provide guidance, and visual guidance can be confused with the original VA elements . Therefore, using a different channel to communicate guidance can help reduce visual complexity and clearly separate the guidance from the original VA system . These are all important for guidance to effectively support user decisions in VA to ensure their trustworthiness.

As previous research in human sensory systems indicates that the tactile channel has the second highest information bandwidth only behind the visual channel [194], an investigation on the potential of tactile feedback to provide guidance is promising. Specifically, as vibration is commonly used and widely available in modern computational devices, I investigated how to provide guidance through vibrotactile feedback, which led to the publication presented in Chapter 10. To summarize its contributions in the context of this thesis, the following first describes the design space of vibrotactile guidance, then presents the results from the user study in two distinct guidance scenarios, and relates the produced insights to trustworthy DM in VA.

5.1.1 *Parameters for Vibrotactile Guidance*

In visualization research, vibrotactile feedback has been mainly studied as a means to encode data, especially for discrete ordinal and categorical data [127, 144, 256]. Meanwhile, vibrotactile guidance has not been formally studied in previous VA research to provide interaction and decision support rather than encoding data. Therefore, I inspected the possible technical parameters that can be tuned to provide guidance through vibrotactile feedback based on previous research. Five parameters – amplitude, frequency, waveform, duration, and pattern – were included in the design space in accordance with a list of widely available computing devices and how they support these parameters. Detailed descriptions of these parameters and the list of devices can be found in Chapter 10.

Individually, these parameters can be used to encode relatively simple information – Humans usually can differentiate between around 4 levels of amplitude [99] and 9 levels of frequency [90]. The difference between commonly used waveforms, such as sine, square, and triangle, is also often easily recognized [98]. Therefore, using these parameters individually is better suited to provide simple decision support in VA, such as reminding user of erroneous actions, or informing users if they have reached an appropriate choice.

Combining these these parameters with various durations and patterns can encode more complex information. Encoding metaphors

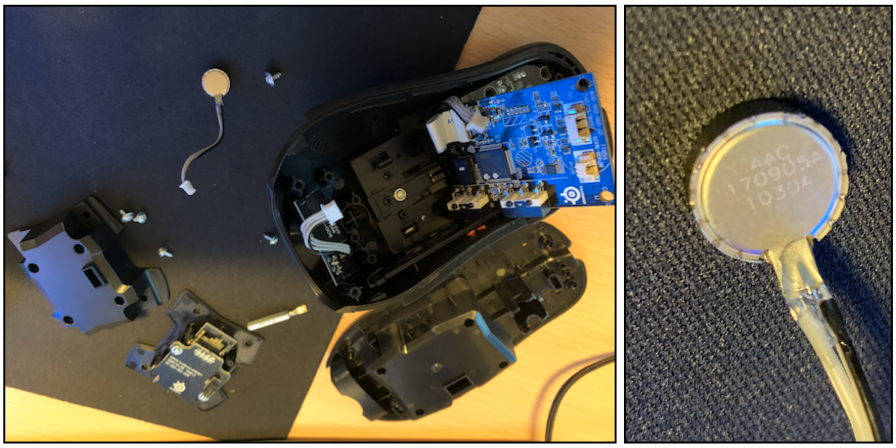


Figure 5.1: The disassembled computer mouse and its vibration motor used in the user study for vibrotactile guidance.

related to its corresponding sensation, such as a small tap, a few knocks, or a buzz, is often easier to understand and learn [5, 28, 29]. Meanwhile, these patterns can also serve as a non-metaphorical language to encode a range of information – the design structure for vibrotactile patterns, Tactons, has been developed to encode different meanings, such as creating and deleting files, or different system errors [25]. However, these patterns usually take time to learn and are more commonly used in accessibility support for people with visual difficulties [208].

5.1.2 *User Study in Two Guidance Scenarios*

To explore how to provide guidance through vibrotactile feedback, experimental prototypes for two distinct guidance scenarios were built using a vibration-enabled commercial computer mouse (see Figure 5.1) with a focus on the amplitude parameter. Two guidance scenarios were built to represent the two types of guidance problems – target unknown and path unknown [41] – in an isolated and artificial context to control possible confounding factors. The selection task corresponds to target unknown guidance scenario, where users aim to use brushing to select certain number of data points in a scatterplot, without explicit knowledge on what is the targeted number of points. The navigation task corresponds to path unknown guidance scenario, where users were instructed to navigate to a specific known data point in a scatterplot, without the knowledge on how to get there. The visual guidance utilized in these tasks are presented in Figure 5.2. Amplitude was chosen as the studied parameter for the vibrotactile cues, since it is best fitted to encode the information of getting closer or further away from a certain target in the two guidance scenarios. Three different vibrotactile cues – increasing, decreasing, and threshold – were developed along with one visual cue as well as three

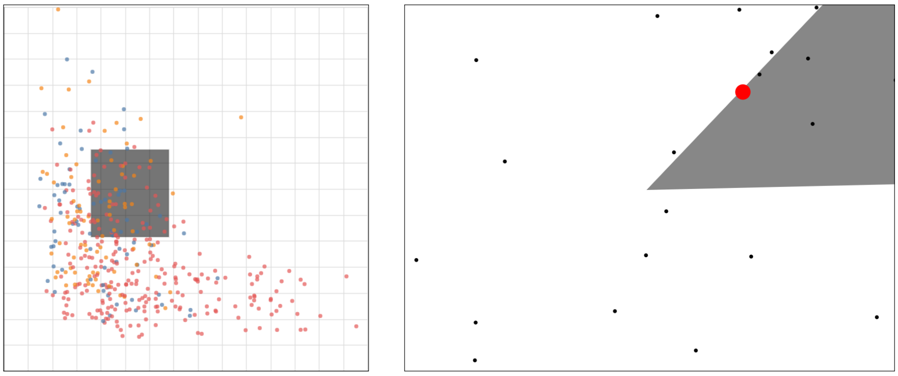


Figure 5.2: The visual guidance used in the user study. The visualization on the left illustrates in selection tasks how close the number of selected points is to the target number with the background of the brush. The visualization on the right illustrates in navigation task if and how fast user is moving towards the target point with the background of the arc.

respective combinations of visual and vibrotactile cues. Increasing cue turns up the vibration amplitude when users select data points closer to the target number or move closer towards the target point, while decreasing cue turns down the vibration amplitude. Threshold cue only triggers the highest vibration amplitude when users select the right number of data points or are moving along the ideal path.

The user study was conducted individually with 14 participants. User performance (time and error for the selection task, and time for navigation task) and user experience (a 7-point Likert scale question and semi-structured interviews) were evaluated for each guidance cue in each task.

For the selection task, while the improvement to visual cue was not significant, threshold vibration did show a better performance in error and yielded the best user performance in both time and error when combined with visual cues. Meanwhile, decreasing vibration combined with visual cue was rated as the best in user experience. For the navigation task, increasing vibration yielded the best user performance and was rated as the best in user experience when combined with visual cue.

These quantitative results were also reflected in the qualitative interviews – For the selection task, participants (P1, P4, and P13) mentioned that the threshold vibration gave them a sense of accuracy. Meanwhile, combining it with the visual cue helped the participants first find the rough area of the right number of points with the visual cue, and then the threshold vibration allowed the participants to pinpoint the exact number of points to select. Moreover, the decreasing vibration was preferred by users as detecting "no vibration" in the decreasing pattern is easier than identifying the "maximum vibration" in the increasing pattern. In the navigation task, however, increasing vibration was preferred as it causes less disruption to users. Participants (P2, P8, and P12) stated that the decreasing pattern would vibrate unless they are on the ideal track, leading the mouse to "vibrate all the time". For the threshold pattern, participants (P1, P9, and P12) found it too hard to trigger.

5.1.3 *Discussion*

Overall, this study explores the possibility of communicating VA guidance through vibrotactile feedback, which helps to reduce potential visual clutter and perceptually separate guidance components from existing VA elements. To establish the design space of vibrotactile guidance, I discussed its possible parameters based on previous research and inspected a list of widely available devices with their capabilities in these parameters according to the references in their technical documentations. This provides both conceptual and practical foundations for VA researchers to further explore the possibilities of vibrotactile guidance.

In the user study, one vibrotactile cue in each task (threshold for selection accuracy, and increasing for navigation time) led to better user performance than the visual cue, although the improvement was not statistically significant. Nevertheless, preliminary insights from the study show that vibrotactile cues could, at the very least, enable user performance and experience as good as visual cue. For the purpose of using an alternative perceptual channel to provide guidance in this study, the results validate that vibrotactile cues could serve as an alternative channel for communicating guidance without significantly deteriorating user performance and experience. Meanwhile, vibrotactile cues could potentially improve user performance compared to visual cue in some cases, while they come with the benefits of reducing visual clutter and separating guidance from existing VA elements. To further validate these benefits that contribute to trustworthy DM, more research is needed in context-specific settings where users are already mentally overloaded with complex VA systems.

Another interesting insight from this study is how the vibrotactile cues work differently for the two tested tasks. For the selection task, threshold vibration performed the best among the vibrotactile cues in accuracy, and participants also commented on the accuracy it offers. When combined with visual cue, it also led to the best results in all measured metrics (accuracy, time, and subjective experience). A participant also mentioned that such a combination provided "a sense of security" by allowing them to first select the approximate number of ideal points with visual cue and pinpoint the exact number with the threshold vibration. This shows how vibrotactile guidance can help support trustworthy DM by providing a sense of confirmation to users. For the navigation task, however, increasing vibration led to the best user performance (lowest completion time). Participants also criticized how decreasing vibration in this task is constantly present, while praised how increasing vibration keeps them informed when they are in the "right direction". This shows how vibrotactile guidance could also disturb users, and users prefer to be confirmed when they are doing the right thing instead of constantly being reminded what they have done wrong. In the context of supporting trustworthy DM where explicit guidance on user decisions is provided, this implies that decision support should not be excessive and might be more effective when its communication focuses on the positive side of confirmation than the negative side of "nudging" users.

In a wider research landscape, the contributions of this work also made the first inquiry into multimodal guidance in VA – Existing guidance research often mentions non-visual channels as potential options for presenting guidance, while application of non-visual guidance is hard to find. The exploration of vibrotactile guidance in this work opens up such potential and has made an impact on further research into the possibilities of other non-visual forms of guidance [145].

5.2 SKETCHY RENDERING

In regular visualizations, the visual presentation is often highly repetitive and similar among different areas. This poses a challenge on users to orient themselves and accurately recall the location of important data points. Although visually marking these data points could help to resolve such challenge [75, 283], regular visualizations might have used colors as an encoding technique to present the information in each data point in the otherwise already complex visual presentation. Therefore, a potential research opportunity lies in utilizing some of the existing elements in regular visualizations to orient users and aid their recollection. Easier and more accurate recollection of such information is also relevant for trustworthy DM, as this could help reduce the cognitive load of users and make room for more mental resources. A more accurate recollection of information also ensures a trustworthy basis which user decisions rely on and increases the trustworthiness of the knowledge that propagates through the DM process.

To this end, this section presents a sketchy rendering technique to add distinctive characteristics to existing grids of regular visualizations, as sketchiness is rarely used to encode data characteristics in visualizations [21]. User study was conducted on a developed prototype to investigate its potential in aiding the recollection of data locations. To summarize its contributions, the following describes the design space for sketchy rendering aiming to add unique characteristics to regular visualizations, presents the results of the user study comparing two sketchy rendering approaches with regular straight-line rendering, and relates the insights to trustworthy DM in VA. Further details of this technique can be found in Chapter 11.

5.2.1 *Parameters for Sketchy Rendering*

Extending existing research on sketchy rendering [35, 281], the proposed design space for sketchy rendering in regular visualizations is composed of two steps – drawing the sketchy lines and composing the sketchy shapes – with a special focus of bringing out distinctive features.

Six parameters were proposed to produce distinctive sketchy lines, with four parameters changing the geometry of the line and two parameters changing the stroke width of the line. Visual examples of these parameters can be found in Chapter 11.

To further compose the sketchy grids with the sketchy lines, two approaches were proposed.

- *Line-based* approach that renders each line across the grid cells individually, and
- *shape-based* approach that renders each cell in the grid individually as a square.

Examples of these two approaches are illustrated in Figure 5.3.

5.2.2 *User Study with Two Rendering Approaches*

The user study aims to uncover how different rendering approaches in regular visualizations compare in their accuracy of transferring perceived information to their knowledge and the speed of retrieving such knowledge. An experimental prototype was developed with two rendering approaches, line-based and shape-based, as the independent variable, while the straight-line rendering without sketchiness was included as a baseline comparison.

I designed and conducted the user study with 16 participants in a controlled environment to compare these two rendering styles and the baseline straight-line rendering. Users were asked to first observe a 16-by-16 grid with a highlighted cell on a randomly generated location in the grid for 5 seconds and later to recall the location of the cell by clicking on the corresponding location after a distraction task. Time and error of the recollection were measured in each rendering

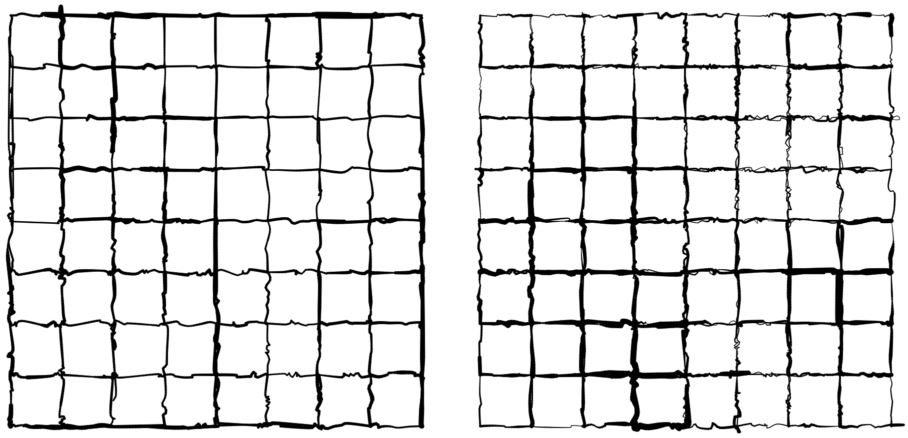


Figure 5.3: Line-based (left) and shape-based (right) grids with $p \in [1..18]$, $a \in [1..5]$, $d \in [\pm 1.. \pm 15]$, $o = [+1]$, $sw \in [1..5]$, $sc = 2$.

style. When the participants correctly recalled the location on the first trial, time was measured and the error is recorded as 0. When the participants did not correctly recall the location on the first trial, the Manhattan distance between the clicked cell and the target cell was measured to reflect the error. A short interview was also conducted for each participant to understand their strategies for memorizing and recollecting the highlighted cell.

Results among all the studied participants did not show a significant difference in recollection time and accuracy when comparing the two sketchy rendering styles with straight-line rendering. However, during the interviews, several participants mentioned that they did not consciously utilize the sketchiness in the grids to recollect the location of the target cell. When applying the same analysis to the subset of participants who did consciously utilize the sketchiness, line-based rendering performed significantly better than straight-line rendering in time, while the error is also lower with line-based rendering than with straight-line rendering but not statistically conclusive. When asked about their strategies for recollecting the location of the target cell, the participants who did consciously utilize the sketchy grids all reported that they specifically focused on the distinctive features around the target cell to memorize and recollect the specific location.

5.2.3 *Discussion*

This study explored the potential of aiding the recollection of information with sketchy rendering in regular visualizations through their distinctive features. The results showed that conscious use of the line-based rendering could help to reduce the time and increase the accuracy of recollection, compared to normal straight-line rendering. This provided early findings on how adapting existing visual elements in VA through alternative techniques could help to reduce users' mental effort and increase the trustworthiness of information recollection. Meanwhile, further studies in the context of real-world VA decisions can help to understand more deeply how this information transfers into higher trustworthiness of human decisions.

A particularly interesting insight from the study was on the users' consciousness of the sketchiness and its utility. In the user study, participants were informed about the utility of the sketchy grids – to help them recollect the location of a target data point. However, considering its application in a real-world system, users can be confused with the look-and-feel of the sketchy rendering without knowing its intention. Even with the utility informed in the user study, 10 out of the 16 participants still did not consciously utilize the sketchy rendering for recollection – this is partially due to the fact that we intentionally did not instruct users with specific strategies for using it. Meanwhile, some participants also noted that they did not think the sketchiness would have helped them with recollecting information. From the perspective of implementing VA systems, it could be helpful to clearly inform the utility of such alternative manner of utilizing existing visual elements in order to promote conscious use of it. Additionally, allowing users to opt in/out of such alternative technique also helps to adapt to different preferences among users.

5.3 SUMMARY & REFLECTION

This chapter presents two alternative techniques along with their design spaces and prototypes that focus on providing decision support in VA without adding visual elements. Reflecting on the components in the trustworthiness of human decisions in Chapter 3, the two alternative techniques in this chapter cover mainly the *perception* and *knowledge* components. The main goal of the two techniques was to provide decision support without adding visual elements through alternative means of encoding, thus putting a focus on the *perception* component of human decisions. Meanwhile, the two techniques also support the *knowledge* component in their own way. The vibrotactile technique provides guidance that informs users' interactions in VA systems, while the sketchy rendering aids users' recollection of information.

Compared with what is typically considered decision support in DSS research, the two techniques in this chapter are also "alternative" in the means of supporting decisions. As mentioned in Chapter 2, DSS are generally structured with language, presentation, problem processing, and knowledge systems to support DM. However, adding such complexity on top of an existing VA system is exactly what the techniques in this chapter aimed to avoid. Instead, vibrotactile guidance provides simple knowledge about user interactions through changing the means of support from visual presentation to vibrotactile feedback, while sketchy rendering implicitly facilitates users with their recollection of data locations through the sketchy features in the visualization grids.

Finally, although the two techniques avoid adding visual elements while providing decision support, the benefit of lowering users' cognitive load is dependent on the assumption of a direct correlation between the introduction of new visual elements and an increase of cognitive load. For vibrotactile guidance, a potential visual element was substituted with a vibrotactile element, which still might bring in a comparable amount of cognitive load. For sketchy rendering, although no new element was added, the perceived complexity of the element could still have increased with the introduction of sketchiness. Therefore, further validation of their usefulness in lowering the cognitive load of users is needed through testing the techniques in complex VA scenarios. Nevertheless, these techniques are still useful to support user decisions in scenarios where visual encoding or space is limited or exhausted.

6 CONCLUSION

Through building the theoretical foundations, developing a generic method, and exploring alternative techniques, this thesis investigated the trustworthiness of DM in VA from different perspectives. In this chapter, I discuss possible directions for future work and conclude this thesis with a summary of the contributions.

6.1 FUTURE WORK

As some of the first research on the trustworthiness of human decisions in VA, this thesis has primarily focused on a more general and abstract level to connect VA with relevant research. To further the understanding of human decisions in VA and ensure their trustworthiness, both empirical work in real-life VA scenarios and intermediate-level knowledge [116] such as frameworks and models are needed. In the following, I outline the potential directions for future research based on the contributions in this thesis.

Support Trustworthy DM in Various VA Contexts

As the generic method in Chapter 4 aimed to provide a general approach to design VA guidance as decision support, it was not formulated with a specific VA system or context in mind. Similarly, user studies with the techniques in Chapter 5 were also carried out in abstract settings to focus on the techniques themselves and exclude potential confounding factors. Meanwhile, as research in data analysis indicated: human decisions during the analytical process are often complex, iterative, and dependent on the contexts [155, 157, 232]. Therefore, understanding human decisions in real-life VA scenarios and building solutions for various contexts are important next steps to continue the research agenda outlined in this thesis. On the one hand, empirical research on how users make decisions in real-life VA scenarios is needed to fully understand how DM theories apply in the context of VA and build a solid foundation for developing techniques to effectively support the trustworthiness of human decisions in VA. On the other hand, coupled with empirical research, development of various approaches for trustworthy DM in different contexts is needed to adaptively and comprehensively support complex human decisions in VA.

Measure the Trustworthiness of Human Decisions in VA

To ensure trustworthy human decisions in real-life VA scenarios, another important challenge is to measure the trustworthiness of human decisions. This is important both for machines to evaluate human decisions and provide support when needed, and for researchers to evaluate different VA techniques in their effectiveness of ensuring trustworthy DM. Such measures are relatively easy to obtain when the decisions are simple and the ideal choices are objectively known. For example, in the selection task for the vibrotactile guidance in Chapter 5, the ideal number of data points to select was preset and known in the system by design, so the trustworthiness of user decisions could be measured by accuracy. However, as indicated in the generic method in Chapter 4, many human decisions in VA are based on multiple criteria, some of which can even be implicit, subjective, and difficult to articulate. Such implicitness and subjectivity of human decisions are also important reasons why this thesis is titled with the word *"trustworthy"* instead of *"accurate"*. While existing research provides means to evaluate the decision alternatives in VA, such as quality metrics of visualizations [12] and evaluations of algorithms [141], evaluation of human decisions is so far limited to preliminary metrics for detecting biases [270]. Further research is needed to understand how humans make decisions in VA and develop general methods for measuring the trustworthiness of human decisions in VA.

Develop Models for Trustworthy DM in VA

In this thesis, the theoretical foundation in Chapter 3 provided pointers to the components of human decisions to consider, while the generic method in Chapter 4 offered clear steps and practical tools to provide guidance that supports human decisions in VA. To further enable researchers to understand the relevant concepts of human decisions in VA and practitioners to develop VA systems that support trustworthy human decisions, universal models distilled from empirical research in various contexts and relevant theories are needed. For guidance design to support DM in VA, a nested model similar to Munzner's nested model for visualization design [187] is within reach through combining the guidance method in this thesis with other guidance frameworks, e.g. Pérez-Messina et al.'s [203] task typology as task abstraction and Sperrle et al.'s [244] Lotse guidance library or MCDA mentioned in this thesis as algorithm design. Beyond the concept of guidance, a general model for trustworthy DM in VA is promising by connecting the components of human decisions in Chapter 3 and Simon's [236] model of DM process with existing VA models such as the Knowledge Generation Model from Sacha et al. [221, 222].

6.2 CONCLUSION

With a growing amount of data collected, processed, and analyzed in every corner of our lives, the trustworthiness of data analysis also becomes increasingly critical for informed, inclusive, and fair decisions that could influence virtually every human being. Meanwhile, supporting DM has been cited as one of the most important utilities and goals of VA since its birth [134, 257]. Subsequent research made important contributions towards supporting users to make better decisions with the help of VA. However, this view of the relationships between VA and DM comes with the critical limitation of overlooking decisions during the analytical process that significantly influence the outcome of the analysis. Furthermore, the concept of trust in VA research has been mostly taken as a one-way street of how human users perceive the trustworthiness of VA systems, while omitting the trustworthiness of human decisions in this mutual relationship.

To address these research challenges, I presented a new perspective on the concept of VA by rethinking its relationship with DM and trust. Shifting the focus from supporting decisions *with* VA to the decisions made *in* VA during the analytical process, this thesis took user tasks in VA systems as DM processes and connected them with relevant research. Taking the trustworthiness of human decisions into consideration, this thesis also drew a more complete picture of the trust dynamics between humans and machines in the analytical process. Towards trustworthy DM in VA, this thesis explored its theoretical foundation, generic method, and alternative technique.

- Through connecting VA with research in DM and trust, the theoretical foundation laid the groundwork for understanding human decisions in VA and their trustworthiness. Specifically, putting relevant DM theories in the context of VA provided means to identify and analyze human decisions, while inspecting the trustworthiness of human decisions outlined possible components that need to be considered and supported in practice.

- By reframing VA guidance as decision support, the generic method provided practical tools to design and develop guidance for supporting trustworthy DM in VA. Following the *Intelligence-Design-Choice* model of DM process from Simon [236], I adapted an interaction taxonomy to structurally identify user decisions in VA that call for guidance, integrated MCDA models as the mechanism for guidance generation to evaluate the alternatives in these decisions, and composed views of multiple alternatives to present guidance as decision support. Through expert workshop and prototype development, the method was realized with a set of worksheets and illustrated with a use case.

- With vibrotactile guidance and sketchy rendering, the alternative techniques explored the possibility of providing decision support in VA without adding visual elements. In the vibrotactile guidance technique, this alternative perceptual channel of VA guidance was studied to support users in confirming their interactions. In the sketchy rendering technique, this alternative way of utilizing existing visual element in VA was tested out to support the accuracy and efficiency of recollecting information. User studies with the prototypes of the techniques showed their potentials and provided empirical insights for supporting trustworthy DM in VA.

In sum, through connecting a wide range of interdisciplinary research, including VA, DM, and HCI, this thesis draws research attention to the trustworthiness of human decisions during VA processes, expands the concept of VA with a fundamental understanding of what trustworthy DM in VA encompasses, and provides potential means to support user decisions in VA.

Part II

PUBLICATIONS & MANUSCRIPTS

7

PAPER A: TRUST CALIBRATION

Beyond Trust Building — Calibrating Trust in Visual Analytics

Wenkai Han and Hans-Jörg Schulz

Department of Computer Science, Aarhus University, Denmark

ABSTRACT

Trust is a fundamental factor in how users engage in interactions with Visual Analytics (VA) systems. While the importance of building trust to this end has been pointed out in research, the aspect that trust can also be misplaced is largely ignored in VA so far. This position paper addresses this aspect by putting trust calibration in focus – i.e., the process of aligning the user's trust with the actual trustworthiness of the VA system. To this end, we present the trust continuum in the context of VA, dissect important trust issues in both VA systems and users, as well as discuss possible approaches that can build and calibrate trust.

7.1 INTRODUCTION

In one of the most cited paper on Visual Analytics (VA) [134], Keim et al. proposed that VA should *integrate scientific disciplines to improve the division of labor between human and machine*. By integrating human expertise through the human-computer interaction, VA systems aim to enable data experts to explore data graphically and generate insights more easily. However, as users grow dependent on the VA systems, new uncertainties and errors that the VA systems bring in might expose users to the risk of generating ill-informed insights. This would be detrimental for VA system – if users become aware of such uncertainties and errors, they might lose their trust in the VA system and stop using it; if users stay blind to the uncertainties and errors, the ill-informed insights they produced might cause them to make problematic decisions. Such issues coincide with previous trust research – trust is increasingly relevant under the conditions of uncertainty presence in the trustee (VA system), vulnerability to risk for the truster (user) and dependence relationship between the truster and the trustee [136].

Previous research on trust in visualization has mostly focused on the idea of trust building – essentially to improve users' trust in VA systems [172, 221]. However, VA systems are designed always by humans and subject to potential human errors and subjectivity. Furthermore, one of the fundamental ideas in VA – human-in-the-loop – emphasizes that humans should supervise and steer the analytical process to generate trustworthy insights. Therefore, it is necessary and positive for users to maintain a healthy skepticism towards the VA system. In this position paper, we consequently propose that calibration of the appropriate trust level is equally important as, if not more than, trust building. With these concepts, we mean concretely:

Trust building increases the trust a user puts in a VA system through various means, such as making computations transparent through visualization (showing *what* the system is doing), providing explanations for results (showing *why* the system is doing it), and allowing the user to interject and reparameterize at any point.

Trust calibration aligns the trust put into a VA system by the user with the system's actual trustworthiness through various means, such as communicating uncertainties, providing visual cues and previews of the end result the user can expect from the system, and indicating analysis paths that have shown to work for similar data in the past.

In Sec. 7.2, we first lay out the trust continuum as a basis for the discourse of trust building and trust calibration. Then, Sec. 7.3 dissects potential trust issues in both VA systems and users and outlines possible approaches to build and calibrate trust. Sec. 7.4 subsequently connects some emerging VA approaches with the previous discussions of trust to inspect how they might bring new perspectives for the trust dynamics in VA. At last, Sec. 7.5 concludes this paper with some overarching insights and recommendations for future research regarding trust calibration for VA.

7.2 CONTINUUM OF TRUST

Trust building and calibration deal with trust issues from different but complementary angles. Trust building emphasizes increasing users' trust level in VA systems, while trust calibration focuses on avoiding and mitigating misplaced levels of trust. This difference is illustrated by the trust continuum shown in Figure 7.1: where trust building aims to increase the trust level from left to right, trust calibration aims to align the trust from bottom to top. The elements of this continuum are introduced below.

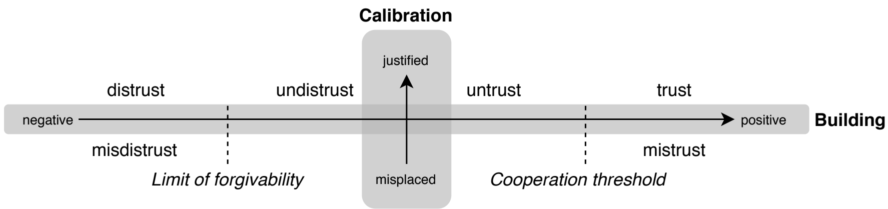


Figure 7.1: The trust continuum extended on the model of Cho et al. [49]

7.2.1 *The Foundation: Trust and Trustworthiness*

The definition of *trust* varies in different contexts, but the general concept of trust is defined as *"the belief that the trustee will act in the best interests of the truster in a given situation."* [169]. This captures the dynamics of the trustee and truster in a social relationship. However, this changes in the context of VA, as one of the two trust parties, the VA system, is a largely non-social actor.

In the situation of VA, the primary goal, i.e. the "best interests" of the truster (user), is to *"identify and visually distill the most valuable and relevant information content."* [134] Therefore, we can adapt the definition of trust in the context of VA as *the truster (user)'s belief that the trustee (VA system) will help them correctly identify and visually distill the most valuable and relevant information content.*

Note that trust is slightly different from *trustworthiness*. While trust is a *belief* that is not necessarily based on observed evidence, trustworthiness is the verified and objective trust based on observations [241]. In the context of VA, we can think of trust as such belief that users might have about the VA system and that is possibly even preconceived and formed without actually ever having used the system. Whereas trustworthiness is based on the observation that the VA system helped users to achieve their goals and the expectation that the VA system will behave consistently in that regard.

7.2.2 *Levels of Trust: Distrust, Untrust, and Undistrust*

In addition to the state of full trust, there are three more levels of trust: distrust, untrust, and undistrust.

Distrust measures an active form of negative trust where the truster believes that the trustee will actively work against their interests [169], which can lead to disuse, i.e. abandonment, of digital systems [150]. When users distrust a VA system, they may have found that the VA system repeatedly produces inaccurate visualizations. While this may not be "malintent" by the system or its authors – e.g., in cases where complex data standards are not fully supported [230] – it can still hinder carrying out an analytic task consistently and free of errors.

In the worst case, this causes users to no longer deem the system trustworthy and thus abandon it. On the brink of distrust, trust building is vital, as users are likely to disuse the VA system.

Untrust, on the other hand, indicates a state where the truster is *not fully confident* in the trustee, while being at the same time still inclined to trust it for the most part. For VA systems, it is natural for users, especially experts, to be alert and consider if there are any errors in the data, implicit assumptions in the computational process, or overplotted information in the visual representation. In particular, when users are not yet fully acquainted with a VA system, such considerations might help them to be aware of the implications of their analytic choices they may not yet be aware of.

Undistrust means the lack of trust[34], where the truster becomes *suspicious* of the VA system, but has not fully distrusted it. Compared to untrust, undistrust leans more towards the negative side, where the truster contemplates more to distrust the trustee. In the state of undistrust, users have serious doubts about the VA system and its trustworthiness, but they can still perform most of their intended tasks and generate some insights when using it with caution.

7.2.3 *Misplaced Trust: Mistrust and Misdistrust*

Mistrust and misdistrust denote situations where trust or mistrust is *misplaced* compared to the trustee's actual trustworthiness. In other words, the level of trust brought forth by the truster and the trustworthiness of the trustee are miscalibrated and the user's expectations of the system do not align with what the system can actually provide.

Deriving from the notion of misinformation, *mistrust* is often called *misplaced trust* [168]. It arises when the truster gives a positive estimation of the trustee that later proves to be misplaced. This is particularly problematic, as mistrust can lead to misuse, i.e., users generating inaccurate results and gaining false insights, which works against their interests of using VA systems. Furthermore, later when users find out such mistakes, it is more possible for them to feel "betrayed" or "cheated" by the VA system and start distrusting it.

Defined as *misplaced distrust*, *misdistrust* is the counterpart of mistrust, where a truster distrusts a trustworthy trustee [173]. Misdistrust originates from miscommunication or misunderstanding between the user and VA system. Misdistrust is detrimental to the interaction, as users might disuse the VA system, when in fact the VA system can be trusted. Once misdistrust has formed, it can eliminate the possibility of the VA system to later "redeem" itself.

Along the same lines, de Visser et al. proposed a trust calibration model between the level of trust and the actual trustworthiness [267].

When the trust level is higher than the actual trustworthiness, they speak of *over-trust*, whereas a lower trust level than the trustworthiness is termed *under-trust*. Note that there is a key difference between over-trust/under-trust and mistrust/misdistrust. Over-trust and under-trust can refer to any situation where the user's trust is higher/lower than the actual trustworthiness of the VA system, even if not by much. This would be the case, when users generally untrust a system that may in fact not be fully trustworthy, but that could still be used with caution – and whose actual trustworthiness is thus on the undistrust level. However, mistrust and misdistrust pinpoint the specific problematic scenarios where users trust or distrust a system that should not be trusted or distrusted, respectively.

7.2.4 *The Bounds of Trust: Cooperation Threshold and Limit of Forgivability*

The concepts of *"cooperation threshold"* and *"limit of forgivability"* were introduced by Marsh and Briggs [168]. They delineate trust and untrust, as well as distrust and undistrust, respectively.

Cooperation threshold refers to the point beyond which trust is established and the two parties will jointly proceed towards the same goal [168]. In a social context, *cooperation* means the action of different people working together, whereas in the context of VA system and the user, we define cooperation as fluent, reliable, and convergent interaction between system and user that work towards jointly identifying and distilling valuable and relevant information. Note that human usage of a VA system alone does not constitute as full cooperation, but that it requires the mutually dependent nature of the iterative human-in-the-loop interaction with each other.

Limit of forgivability refers to the limit beyond which the trustee is truly distrusted and can be considered only as acting against the truster's best interests. According to Marsha and Briggs [168], this limit determines the worth of the trustee entering into redemption strategies to seek forgiveness from the truster. In the context of VA, we can see this as the limit beyond which a deeply disappointed user would abandon and disuse a VA system.

7.3 SHOULD I TRUST, AND WHAT TO TRUST?

VA provides users with powerful tools for understanding and reasoning. However, VA systems also confront users with computed results and mined patterns that stand in conflict with the user's previous knowledge, experiences, and beliefs. This leads to a series of questions: "Should I trust myself or the VA system?", "How much should I trust the VA system?", and "Which part of the VA system should I trust more?" To answer these questions, analysts must know about the strengths and the weaknesses of both sides – the VA system and themselves – to know whom to trust in which situation. Thus, the following dissects potential trust issues on both sides, provides pointers to existing research for each, and details what can be done to build and calibrate trust in each case.

7.3.1 *Should I Trust the VA System?*

VA systems are designed by humans and therefore subject to potential human errors and subjectivity. Moreover, VA systems rarely have access to the "big picture" of the context behind a given analytic task. For example, a VA system does not know that reporting a computed result to the 10th digit after the comma miscommunicates a level of certainty and detail that is not warranted when averaging 5 roughly estimated numbers, leading to mistrust/over-trust in that result. In this section, we dissect how trust issues emerge in different parts of VA systems – data, computational process, visualization, and interaction – as well as what can be done to address these issues.

7.3.1.1 *Should I trust the data?*

"Garbage in, garbage out." This principle captures the observation that the quality of the input to a digital system is directly reflected in the quality of the produced output. It also holds true for trust in VA: if the input data to a VA system are not trustworthy, then this lack of trustworthiness will propagate all the way to the derived insights.

As much as people label datasets as "raw", such data are still collected through certain technical and social lenses. The "raw data" we obtained "are always already cooked and never entirely raw" [91], and thus raise questions of trust. National population census data in some countries are collected through investigators going into every household and might be subject to various human errors. Natural sciences researchers place sensors with varying accuracy in locations that they deem as reasonable to gather data for their research. Tech companies collect user data through their own algorithms, selecting data that are relevant to their field of business, easy to access, and legal to be collected. As such, even the data in their most original forms are conceived before the collection process and limited by various technical and social constraints. When such conceptions and constraints are not communicated to the users of the data, inconsistencies in the "raw data" can be easily overlooked and lead to mistrust, or even be misconstrued as intentional manipulations and lead to misdistrust.

In a review paper on trust in digital information, Kelton et al. concluded that people tend to put more trust in accurate, up-to-date, complete information without deception and distortion, and is persistently obtainable with responsible methodology [136]. Therefore, to calibrate the trust to be placed on the input data, the inclusion of related information about the data source and communication of uncertainties in the collected data are essential [230]. Such metadata can inform users about where data discrepancies stem from and make users aware of the impact these discrepancies have on their analysis. Metadata make the processes transparent by which the data were gathered and further processed. However, such metadata are not proof of this process being the most suitable and they rarely explain why a particular process was chosen. Adding this reasoning behind them would further help to judge the data's trustworthiness.

Yet we also need to communicate the metadata to the user to make a judgment of trust. Uncertainty visualization is a frequently mentioned approach to communicate quantitative uncertainties [19]. In theory, communicating such metadata should allow for better judgment of the data and thus of any processing result based on that data. In practice, though, it turns out that most users have a hard time to reason with uncertainties, let alone to parse the provided visualizations [118]. As for qualitative uncertainties originating from the process of data gathering and preprocessing, communicating the data provenance is an established approach [110]. Given the data provenance reflects a systematic and responsible methodology behind it, it has the potential to instill trust in users. In addition, such openness about the process behind the data can give an impression of "we have nothing to hide" and increase the trust level in general.

7.3.1.2 *Should I trust the computational process?*

The computational process in VA systems is like a black box ingesting data and producing results to be subsequently visualized. As such, it provides little to no internal status to understand its inner workings. Having little insight in and understanding of the computational process, it is almost inevitable for the users to start assuming "intents" of a VA system – likely negative ones. Harboring such assumptions, users will actively look for instances where the system appears to work against them, which will eventually lead to distrust.

Many interactive visualization tools emphasize their integration with computational software such as MATLAB and R. However, as Mühlbacher et al. pointed out, such computational software is usually used as a black box that runs in isolation, providing no output other than the final result once it is ready and defeating the purpose of a visual-interactive data analysis [184]. More importantly, users have very limited knowledge of what is going on in the algorithms behind the scenes and limited agency over the process. When errors arise, users rarely have the option to probe into the computational processes to inspect the potential causes, therefore being unable to verify what went wrong and calibrate their trust level accordingly. A user trust study in intelligent systems by Holliday et al. found a similar pattern that without explanations of how the systems work, user trust might deteriorate over time, which is why the perceived transparency of the system becomes increasingly important for users to trust it [114]. Based on currently available computation infrastructure, Mühlbacher et al. subsequently proposed four different strategies to achieve user involvement [184], which in turn provide knowledge about the algorithms, insight into how they run, as well as agency to users to calibrate their trust levels.

In addition to user involvement and understanding, Friedman and Nissenbaum pointed out that technical and social constraints can transfer into issues in computer systems [83]. In the context of computational processes in VA, algorithmic bias is a notable issue. Algorithmic bias touches on systematic errors in the algorithms that might create unfair results. Danks and London gave some good examples on such issues – the training data might be skewed due to moral or legal reasons, or the algorithm could be designed to counter overfitting noisy data but then ending up more biased in other scenarios [58]. If results from such biased algorithms are still consistent with the users' expectations, they might end up mistrusting an actually untrustworthy computational process. It is therefore important to at least identify and communicate potential computational bias from the algorithms to calibrate trust. To cope with algorithmic bias, Cabrera et al. developed FAIRVIS [33] to aid discovering intersectional bias in machine learning and creating more equitable algorithmic systems. Such tools can be helpful to uncover and communicate algorithmic biases, helping to avoid mistrust.

7.3.1.3 *Should I trust the visualization?*

Visualization displays the results from the computational process to make it easier for the human user to gain insights. To do so, most VA systems provide a limited selection of different visual mapping and rendering techniques, and such techniques are very often not an accurate one-to-one mapping from the data space to the view space. While this is only natural in the age of big data where we have many more data points to plot than available pixels on our screens, it still misconstrues the data and is thus a potential cause of distrust.

Many visualizations are visually pleasing, which can help to build initial trust, especially with inexperienced users. However, if such visually pleasing graphs do not communicate the underlying data accurately and provide effective means to discover insights, such initial trust will sooner or later prove to be mistrust and eventually lead to distrust. It is therefore important to calibrate trust through providing some form of guidance that can help to avoid mistrust. Recommendation systems such as Tableau's "Show Me" [162] and Moritz et al.'s Draco [183] can to some extent avoid "visualization design mirages" [178] by incorporating design knowledge and guidelines in their recommendations. Furthermore, visualization linting can help to uncover improper visual mappings. Similar to code linting, visualization linting searches for common visualization mistakes and automatically highlights them to help users recognize and potentially correct them [177].

The rendering of visualizations can also be an important trust factor. On one hand, technical constraints like low resolution and inadequate contrast might make it hard for users to clearly perceive the visualization, hindering them from gaining accurate insights [26]. On the other hand, some rendering techniques simply struggle to put all information in the available display space, which can lead to important information being hidden at subpixel resolution. To nevertheless point the user towards this information, Luboschik et al. have shown guidance to be a valuable means [160]. As they highlight display regions in which data at subpixel level deviate from the currently shown view, the VA system is transparent about its rendering limitations and users know where to zoom-in to find any deviations. This transparency aligns expectations and thus actively calibrates the trust in the VA system.

7.3.1.4 *Should I trust the interaction?*

Usability and user experience of the interaction with a VA system are important for the trustworthiness of it. Coherence is especially crucial for users to understand and trust the VA system, as discrepancies in the interaction might trigger users to scrutinize a digital system further [217]. When users take actions in a VA system, they have conscious or subconscious expectations of the system's reactions. Discrepancies between these expectations and the provided reaction pose a threat to a VA system's trustworthiness. A framework that reflects different forms of such mismatch is Tominski and Schumann's conceptual separations, spatial separations, and temporal separations regarding interaction costs with a VA system [258].

Conceptual separations concern the misalignment between the mental model that users have about the system, the implementation model the system adheres to, and the presented model of its interface. If the users' mental model does not match with the presented model, they might subsequently internalize such mismatch as an error in the system, pushing users to scrutinize and even distrust.

Spatial separations relate to the spatial placements and distances between different interactive elements and system reactions. This is problematic when the user's interaction and visual response from the system are inconsistent. Such inconsistency between users' spatial expectations and the actual spatial separations in the interface would make it harder for users to understand the action-effect causality of the VA system, causing confusion at best and misdistrust at worst.

Temporal separations reflect the latency between a user's action and the system's visual response. Users might have some expectations of the duration of certain internal processes, and when the actual latency drastically deviates from their expectations, they will become suspicious of the system and underlying process.

For example, in coordinated multiple views, users' actions in one view are expected to influence several others. Yet, if this influence is not clearly represented across the different views (spatial separation), or the actions take too long to propagate to other views (temporal separation), users might not be able to understand which of their actions impacted in which ways the other views and misinterpret the underlying logic (conceptual separation), leading to misdistrust.

A way to counter such trust miscalibration is to communicate the system's response and latency regarding users' possible interactions. For conceptual separations, scented widgets [277] can serve as a preview to align users' expectations of their actions with the reactions from the systems by adding cues to the corresponding interactive elements. For large spatial separation between users' action and systems' reaction, visual links [252], arrows or highlights can enable users to follow the action-effect causality and instill trust. Regarding temporal separations, providing estimates of computationally intensive actions in either textual or visual forms can help to calibrate user expectation and trust. However, inaccurate estimates can also create even more discrepancies and induce distrust.

7.3.2 *Should I Trust Myself?*

When users interact with VA systems, variations in their perception, knowledge, judgment, and situational state influence their actions. These factors are essential for trust calibration: On one hand, they may be the reason users misplace their trust or distrust in the first place – e.g., because of their confirmation bias, users trust results more if these align with their beliefs. On the other hand, they can interfere when trying to communicate uncertainties or algorithmic details – e.g., when change blindness makes it hard to follow computational updates. Hence, in this section, we dissect how these human factors are related to trust calibration and which means have been proposed to alleviate the issues they cause.

7.3.2.1 *Should I trust my perception?*

Perceptual factors, such as visual expectation, visual memory as well as visual attention, are important in visualization, as it is the foundation of human sensemaking from large and often complex datasets. To calibrate trust, we need to consider if one can perceive visual information true to what the VA system present. To this end, not only should we be aware that there can be a spectrum of perceptual abilities among users, for example, different degrees and types of colorblindness, dyslexia, or autism. We also need to consider that human perception is far from optimal and error-free, as it is evident by the broad range of visual illusions.

Among the perceptual abilities, visual abilities are relatively well-researched. Thus, it is well-known that sensitivity to color deteriorates with age and colorblindness can also seriously limit the quantity and quality of information we can extract from visual representations [234]. Many tools have thus put color perception into consideration to make sure one can trust what one perceives. For example, ColorBrewer specifically enables choosing only colorblind-safe color scales [105]. VisCheck and Daltonize show how a visualization or user interface looks like for users with different kinds of colorblindness and provide corrections [68]. Other perceptual differences like synesthesia or dyslexia have not been the focus of dedicated studies in visualization. However, they can be expected to also impact the perceptual process in VA systems, as underlined by recent work on a "synesthetic color palette" [218].

Regardless of individual predisposition, perceptual errors such as change blindness or line width and sine illusions arise for any user perceiving visualizations. To achieve trustful visual communication between the VA system and the user, they thus need to be considered. Change blindness occurs when people do not notice changes in visible elements of a scene. In the context of VA, users might not be aware of how an animated visualization changes or how a static visualization updates. This in turn makes it hard to judge the trustworthiness of the system with up-to-date information. Nowell et al. discussed possible solutions using morphing, crossfading, and wireframes to draw attention to regions of change in the view space [195]. In addition, due to human's tendency to perceive distance between curves as the minimal distance rather than the vertical distance, line width and sine illusions are widely discussed in statistical graphics literature, especially when representing areas between two curves. Hofmann and Vendettuoli proposed *Common Angle Plots* to address line width illusion [113], and VanderPlas and Hofmann demonstrated possible solutions to counter the sine illusion [266].

7.3.2.2 *Should I trust my knowledge?*

Running an analysis with a VA system, users internalize the resulting information they yield from the system through their existing construct of knowledge. Users with different expert knowledge understand and interact with VA systems differently [185]. In particular, a lack of knowledge leads to uninformed actions that potentially cause misunderstanding of the systems and miscalibration of trust.

Domain knowledge about the analyzed dataset is vital for sensemaking in VA systems. The human sensemaking process is based on framing the data presented by the VA system with their existing knowledge construct. As Klein et al. pointed out, "sensemaking is a process of framing and reframing", that fits presented data into the analysts' knowledge construct [139]. On one hand, users are inclined to trust data that fits with their framing and distrust one that does not, which can lead to mistrust and misdistrust. On the other hand, trust calibration is also weakened when users do not have enough domain knowledge to judge if they should trust the outputs from the VA system or their own framing. Implementing a form of Analysis of Competing Hypotheses (ACH) can be helpful for mitigating such issues of data-specific domain knowledge. ACH refers to an analytical process to aid decision-making regarding issues with different alternative explanations or conclusions [111]. Enabling users to explore several analytical paths can help to validate different framing of the data and ensure a calibrated level of trust.

In addition to domain knowledge, users' knowledge, or expertise level about VA and the specific VA system has an important impact on the users' ability to take appropriate analytical actions. Lack of knowledge about different computational processes might leave them in a trial-and-error mode when aiming to choose one that is consistent with their intentions; insufficient navigation skills around the VA system might make it increasingly hard for users to discover different options and views that would help them to generate more insights; inexperienced users might not be able to spot errors and understand issues arising in the system and take actions accordingly. These issues hinder a smooth interaction with the VA system, which impedes the users' perception of the system being truthful and their trust in their own actions. To mitigate the lack of domain and VA knowledge, knowledge-assisted visualization has been proposed to help users navigate through different methods, parameters, and visualization techniques. For example, Jänicke and Scheuermann built a knowledge-assisted visualization for time-dependent multivariate flow datasets, in which users can store process knowledge to aid later analyses [123]. Their user study also shows that knowledge can be extracted and transferred to novice users with this approach.

7.3.2.3 *Should I trust my judgment?*

Human decision-making underlies errors and differences in judgment. For example, we tend to seek meaning in things and interpret things within our own experiences, often seeing patterns where there are none. This phenomenon is called *apophenia*, and being aware of it and working actively against it is a skill that is hard to come by [142]. Furthermore, facing the same VA system, different people make different judgments, as they look at the system through their own lenses of reality with different personal traits, habits, and behavioral patterns. Such deviations can interfere with trust calibration.

As part of one's subjective construct of reality, cognitive bias is a systematic deviation from rational judgment caused by the use of heuristics in decision making [115]. A taxonomy of cognitive biases for information visualization by Dimara et al. lists and classifies 154 cognitive biases [66]. For example, *confirmation bias* will have users subconsciously look for evidence that is in line with any prior assumptions, while ignoring findings that contradict their assumptions [196]. Seeing a lot of confirming evidence in a VA system can lead to mistrusting it. This is a clear miscalibration, as the system always shows the full story, but the user only pays attention to one side of it. To address selection bias, Gotz et al. [96] showed the similarity of a selected data subset to the full dataset to ensure the selection is representative. Dimara et al. [65] highlighted optimal choices and altered task framing to mitigate attraction effects. Wall et al. proposed real-time metrics to detect bias [270] and outlined a design space for mitigating bias in VA systems [271].

Moreover, many personal factors, such as culture, gender, and personality can heavily influence how one absorbs information and evaluate trustworthiness. For example, people from cultures with high uncertainty avoidance, such as Greece, Portugal, and Poland, tend to make unnecessarily conservative evaluations, while members of low uncertainty avoidance cultures, such as Singapore, Hong Kong, and Sweden, are more prone to take risky actions [84], which might include trusting something more easily. Regarding personality, locus of control (LOC), which measures the degree to which one feels in control of or controlled by external events [205], is one of the more well-studied personality traits. In the context of VA, Ziemkiewicz et al. found that users with external LOC are able to efficiently complete VA tasks even with unfamiliar visualizations like an inclusion hierarchy, while internal LOC users struggle to do so if not presented with a familiar node-link drawing [289]. To adapt to different personal traits, personalized visualization offers a way to create interfaces that cater to the diversities among users [199], which better align with what different users expect and need, and are thus less likely to lead to miscalibrated trust.

7.3.2.4 *Should I trust my situational state?*

When users perform a data analysis, their situational state is an underlying factor to their decision-making. Their current internal mood as well as external environment can influence how they perceive, understand, and take actions. Thus, the users' situational state also affects the trustworthiness of their actions and insights.

Regarding mood, studies in decision-making show that negative moods are associated with exploration rather than exploitation behaviors, as well as introducing changes rather than maintaining the status quo [226], among others [209]. This also impacts the use of VA systems. For example, using VA systems with a negative mood makes the user more likely to introduce unnecessary changes and to explore the visualization for an extended amount of time.

Environmental factors include physical as well as mental ones. Although VA tasks are usually performed in a consistent indoor environment, it not necessarily optimal – especially social factors such as shared offices, interruptions and distractions can make a focused, in-depth analysis session almost impossible. This leads in turn to an increase in perceptual errors, and an inattentiveness to one's own biases and algorithmic biases alike. External mental factors can also have an impact. Risk is an important mental factor among other external situational variables for regulating trust behavior. When users need to make a high-risk decision, they tend to rely on more trustworthy cues and tools [247]. Thus, for a "high-stake analysis" whose outcome will be of great influence – e.g., a trader's investment decision for a fund or a clinician's treatment decision for a patient – analysts are likely to choose methods they have more knowledge about and feel safer with. But this also makes them more likely to misuse methods that do not fit the problem at hand.

Tracking physical parameters like eye gaze and electroencephalogram has been proven useful in gauging users' internal situational factors like mood and intentions [238]. VA systems can also ask questions about users' intentions or external situations before they enter the analysis process. Such information can be used to adapt the VA system to the situational factors.

7.4 EMERGING FACES OF TRUST

Over the past years, a number of different "flavors of VA" have emerged that introduce new possibilities to the generic VA process that goes back and forth between human and computer. As these emerging approaches have implications on trust building and calibration in VA, we briefly discuss three of them in the following.

Progressive VA (PVA) carries out analytic computations in a stepwise manner on data subsets (so-called *chunks*) in order to visualize and interact with partial results already before the full computation is finished [4]. Researchers list building trust as one of the biggest benefits of adopting PVA, as by communicating the progression of the underlying process, users' gain an understanding of how results are generated. It thus enables building and even calibrating users' trust in the computational process [179]. Yet it also opens the question of how much can one trust the shown intermediate partial result [3]? Since this question is very hard to answer, Jo et al. developed a different approach to put a user's mind at rest: their PVA system ProReveal [130] incorporates safeguards that can be attached to a running progressive computation and formulate a hypothesis about the computation result as a conditional expression. As long as this conditional holds true, the user can move on with the analysis process – but the moment it is no longer true (e.g., because new data has meanwhile been processed that contradicts the hypothesis) the user is notified. Their user study validates that safeguards can alleviate the unsure feelings users have about early and intermediate results. This makes for a very interesting case of temporal separation (cf. Sec. 7.3.1.4) where one can only fully trust a result once the computation is completed. Yet making use of PVA's inherent ability to continue the analysis already from a good enough partial result, this point of full trust still lies in the future. Safeguard effectively resolve this separation, as they allow moving on with the analysis, even while still not fully trusting the partial result.

Mixed Initiative VA extends VA into a discourse where human and computer are more on par with each other. To that end, a Mixed Initiative VA system infers the users' potential intentions and likely analytic goal from their interactions with the system, so as to proactively support these intentions and goals. This support can range from automatically setting suitable defaults for parameters, to the system offering guidance on how to achieve those analytic goals [54]. An empirical study by Dasgupta et al. found that for complex sensemaking tasks, Mixed Initiative VA systems can inspire greater trust [59]. In addition to building more trust, Mixed Initiative VA also provides useful tools to calibrate users' trust in themselves and the VA systems. By learning about the intentions of the users, VA systems can adapt to better meet the inferred expectations and needs of the users [166], which Sperrle et al. defined as a co-adaptive guidance process [246]. This is essentially a communication process between VA systems and users to calibrate trust. Relating this dynamic to the trust continuum, Mixed Initiative VA opens up a different direction of trust from VA system to user: the VA system becomes the truster, the user becomes the trustee, and the VA system has to trust the users know what they are doing and behave rationally in order to correctly infer their intentions. Yet as the human user is often irrational, one can already see how miscalibration of trust in the user is a huge challenge in Mixed Initiative VA.

Collaborative VA extends VA from one to multiple analysts, potentially with different backgrounds and expertise, performing the analysis together [108]. Collaboration has been proven to be useful in bringing in diverse perspectives and mitigate individual's limitation of knowledge and cognitive bias. By communicating knowledge, experience and different perspectives between each other, users will be exposed to more new ideas and are therefore more likely to break their habitual behaviors. This can help users to gain a more comprehensive understanding of the information in the VA systems, and therefore calibrate their trust level. Billman et al. conducted a series of empirical studies on collaborative intelligence analysis [17]. They found a reduction in confirmation bias for heterogeneous groups of people with diverse beliefs when using collaborative systems. However, for homogeneous groups with similar beliefs, their initial biases were accentuated. Therefore, to ensure trustful decisions, it is important to promote collaboration with heterogeneous groups of users to make sure that diverse opinions and inputs will be considered.

7.5 CONCLUSIONS

While research dealing explicitly with trust building has been rare in the field of VA, work that emphasizes trust calibration in VA is even rarer. Inspired by work in related research fields such as automation [112, 150, 267] and intelligent systems [114], we make a clear distinction between trust building and trust calibration, and bring attention to the latter for matching users' perceived trust and the actual trustworthiness of VA systems. Admittedly, trust building is essential to avoid distrust situations where users might abandon the VA systems. However, building trust that is higher than the actual trustworthiness of the VA systems might set user expectations too high, leading users to blindly trust the system, which will result in disappointment sooner or later. This is precisely the point of trust calibration, which aims to find the appropriate trust level for a VA system and dataset at hand. Trust calibration can in most instances be understood as a form of communication between system and human user in which expectations are aligned to avoid disappointments.

In this paper, we established the importance of trust calibration through the conceptual space of a trust continuum and discussed it for VA systems and users. However, much more research needs to be done to gain a more comprehensive understanding of trust calibration in VA. To begin with, trust building and calibration can stand in conflict with each other when the actual trustworthiness of a VA system is low, and thus building perceived trust would actively miscalibrate it. Therefore, it is important to consider and investigate how trust building and calibration should coexist. Furthermore, although there has been some research on evaluating trust levels [59, 114], tracking trust calibration can be a dynamic process that requires continuous monitoring of trust and trustworthiness. How to evaluate trust calibration is therefore an important but complicated question to address. Last but not least, as new VA approaches emerge, trust calibration can become more intricate – PVA brings up additional trust issues when working with incomplete results, Mixed Initiative VA starts to asks about the trustworthiness of users, and collaborative VA introduces interpersonal trust to the VA process. Both theoretical and empirical research is needed to fully dissect and investigate the trust dynamics in corresponding VA approaches.

8 PAPER B: DECISION MAKING

Making and Trusting Decisions in Visual Analytics

Wenkai Han

Department of Computer Science, Aarhus University, Denmark

ABSTRACT

Decision making and trust have both become rising topics in the research community of Visual Analytics (VA). Many efforts have been made to understand and facilitate making decisions with VA, as well as build and calibrate trust. However, previous research largely took VA as a tool to facilitate decision making, but did not explore the possibility to dissect each analytical step in VA as decision making and discuss how decision making theories can be utilized to improve the trustworthiness of decisions in VA. Therefore, this paper instead proposes such alternative take on the relation between decision making and VA, inspects the processes of visually analyzing data as decision making, and discusses how to leverage decision making theories to facilitate trustworthy decision making in VA.

8.1 INTRODUCTION

Over the years, a large amount of research has focused on the pitfalls human might make in the visual analytical process. For example, humans are subject to change blindness where they do not notice visible changes in a scene [195], and cognitive bias such as confirmation bias allows people to focus on information that agrees with their preconceptions [196]. In particular, our paper in TREX workshop last year from comprehensively concluded how one not only should be skeptical about the trustworthiness of VA systems, but also need to calibrate their trust in one's own perception, knowledge, judgment, and situational state in order to make the right decision in VA [103].

Assisting decision making has also been seen as one of the fundamental goals of VA system since its birth. In 2008, when Keim et al. set the stage for VA, they pointed out that VA is to help people "ultimately make better decisions", and "state-of-the-art concepts of representation, perception, interaction and decision-making need to be applied and extended" for VA research [134]. This is also echoed in subsequent VA research, where making decisions with VA is often seen as the center piece and ultimate goal of using VA [67]. Recently, we have also seen some attempts of leveraging decision making theories to assist decision making in VA – FairVis from Ahn and Lin focused on identifying the biases in Machine Learning to promote fairer decision making [33]; Cho et al. investigated the anchoring effect and its implications on decision making with VA [48]; Padilla et al. presented a cognitive framework for decision making with visualizations [200]. These efforts all reveal some important underlying issues and propose means or frameworks to mitigate such pitfalls. Such efforts can also be seen as strategies to improve the quality therefore the trustworthiness of users' decisions with VA. However, if we take a closer look at the VA process – from selecting the data and algorithms to calibrating the parameters and visual layouts – in each step of the way, users need to identify the alternatives to choose from and gather information to make a choice between these alternatives. This constitutes a "decision making process". Therefore, we argue that each task users undertake in the VA system can also be seen as a form of decision making, and the process of making and trusting these decisions *in* the VA system is consequential for analysts to make and trust their final decision *with* the VA system.

To further clarify – making decisions *with* VA systems focuses on the final decision supported by VA – such as diagnosing a patient, choosing a stock portfolio or making political decisions, while making decisions *in* VA systems emphasizes each analytical decision in VA that leads to and supports the final decision – which area of the data to zoom in, how to transform and analyze the dataset, what visual encodings should be applied, etc. Analyzing the decisions made *in* VA systems is crucial as these decisions heavily influence but are markedly different from the final decision supported *with* the VA system. In this paper, we offer such alternative perspective on making and trusting decisions in VA – by taking each task and step in the VA system as a decision making process.

In Sec. 8.2, we introduce decision making theories regarding making the choice between different alternatives and discuss how they can help to make trustworthy decisions in VA, specifically compensatory and non-compensatory strategies. Then, Sec. 8.3 relates these strategies to bounded rationality and dual process theories to highlight how they can be leveraged in VA decisions. Subsequently, in Sec. 8.4 we reflect on how theories in decision analysis can be applied for making trustworthy decisions in VA. Finally, we conclude this paper in Sec. 8.5 by extracting some important takeaways and future research pathways regarding making and trusting decisions in

VA. The structure of the mentioned theories in this paper can be seen in Figure. 8.1.

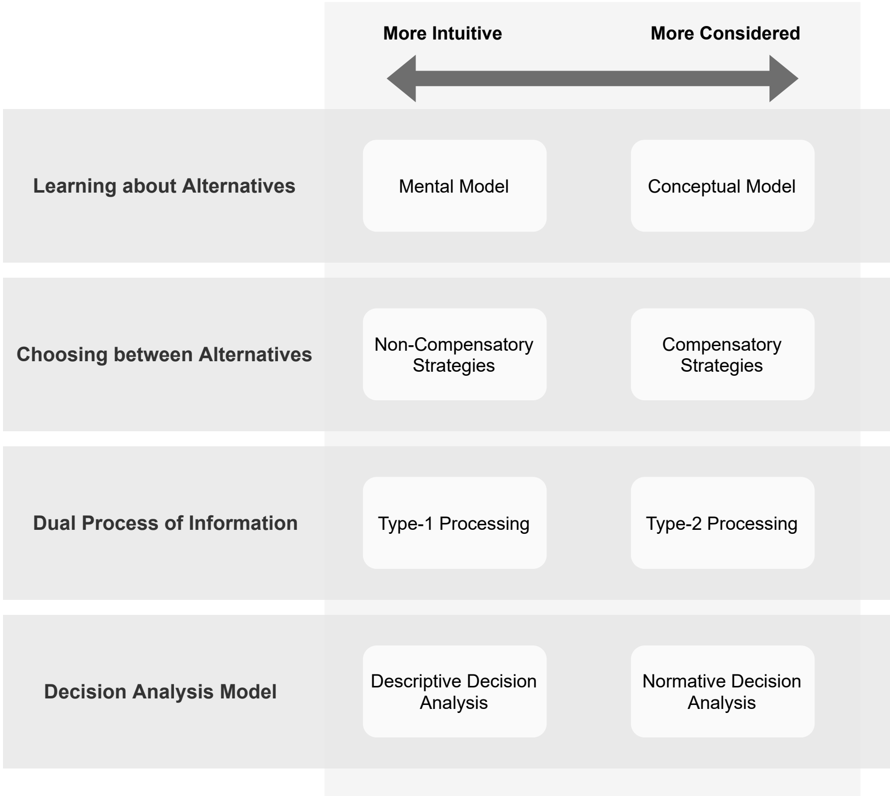


Figure 8.1: The structure of decision making theories mentioned in this paper.

8.2 CHOOSING BETWEEN THE ALTERNATIVES

The central part of making a decision is to come up with alternatives to choose from and make a choice between these alternatives. In VA, many decisions are also done through choosing between alternatives, although sometimes in a more implicit way than making decisions with VA. For example, when analysts choose to focus on one part of the data, they are essentially choosing this subset of data against all the other subsets; when a clustering algorithm is chosen, a decision is made against other clustering algorithms; when a type of visual encoding is applied, analysts also implicitly decided that such encoding is more useful for their purpose than others. In short, each analytical action in the VA process, although sometimes not explicitly framed as a decision, can be always seen as a decision against other potential alternatives. Therefore, the strategies to choose between alternatives are fundamental to be analyzed to understand these analytical decisions.

8.2.1 *Learning about the Alternatives*

To make a good decision, decision makers need to first discover and collect information regarding the alternatives and how they work. In decision making, discovering information refers to the process of identifying a set of valid indicators that might predict the outcome of the decisions. It involves the process to learn about where to look for information regarding the alternatives that later the decision maker acquires and combines to make the decisions. [192] Such process relates to observing how different factors might influence the outcome through "lens of cues" that divides how real world works and how these factors are processed psychologically in a human's mind [102]. In Human Computer Interaction research, this is famously coined as the products' "conceptual model" and users' "mental model" [193]. Both theories assert that how things actually work and how one thinks they work might widely differ. Taking these ideas to the realm of VA – the smallest decision on data, algorithms and visualizations can also produce drastically different results, but the correlation between these factors and the yielded results can indirect and obscure, especially for novice users – a change in the inclusion of a few data points, a tweak on the parameters of an algorithm, or a modification for the specification of a visualization layout all could lead to radically different results. Without understanding the underlying mechanism, users can only make causal inferences about how these factors influence the outcomes.

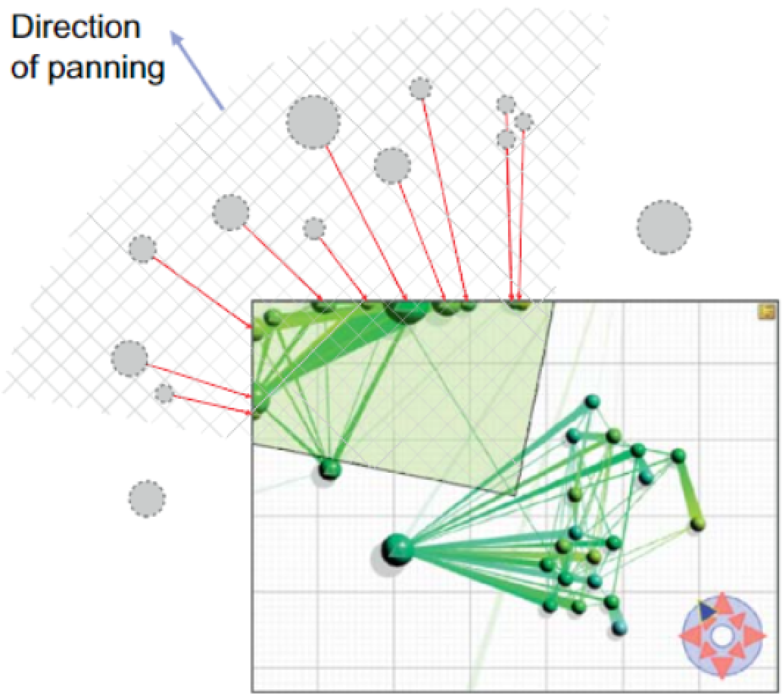


Figure 8.2: The look-ahead radar view [259] uses an arc to indicate direction in which potentially interesting items lie.

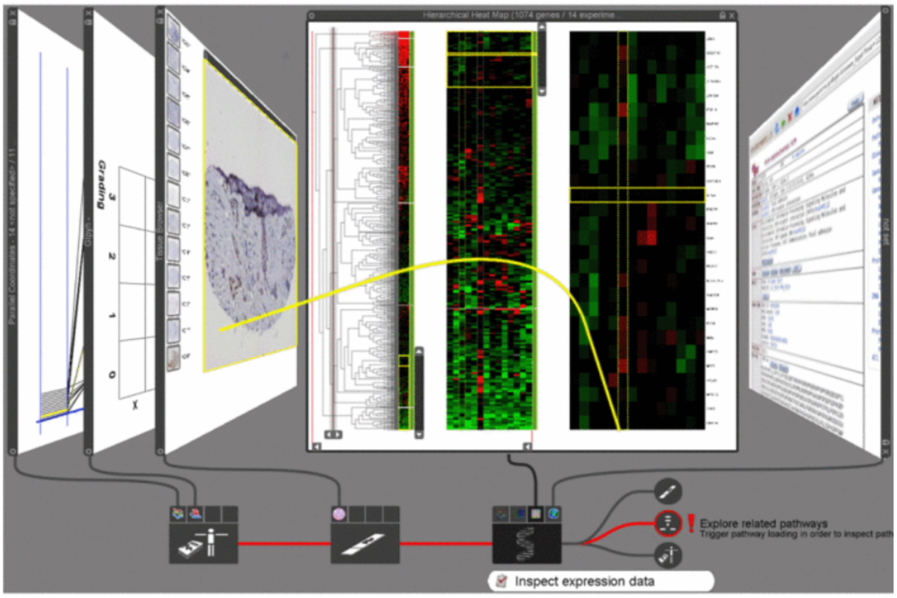


Figure 8.3: The Stack'n'Flip application [255] integrates the data with a map of analytical workflow to present the previous steps as well as recommend future steps to take for users.

Fortunately, in VA systems, there are usually means and resources that users can rely on to understand the system. On the one hand, designers of VA systems often more readily understand the underlying mechanism, and could design the system in a way that guide users towards the useful information. Ceneda et al. characterized the concept of "guidance" in VA as means to resolve a "knowledge gap" encountered by users to execute their tasks [41, 231]. For example, Tominski et al. designed a look-ahead radar view – an arc will appear when users are panning a graph visualization in the direction in which potentially interesting items lie (see Figure. 8.2) [259]. Streit et al. provided a guided view on users' analysis path taken as well as potential future steps that could be taken (see Figure. 8.3) [255]. We also previously explored the potential of using vibrotactile feedback as guidance for users' interactions, where we used vibrotactile cues to guide users to select certain number of data points or find a specific data point in a scatter-plot [104]. On the other hand, decision makers also might have knowledge about the data, algorithms and visualizations that could help them to know where to look at. Therefore, leveraging such knowledge to help users understand the underlying mechanisms and guide users towards important information regarding what to consider could greatly help to produce trustworthy results. VA systems should also reveal the provenance as well as important relevant information of their decisions to help users better judge the trustworthiness of the alternatives suggested by the VA systems.

8.2.2 *Compensatory and Non-compensatory Strategies*

To make a good decision, analysts unavoidably need to choose from a range of alternatives. In fact, one of the most important techniques to improve decision making is to "adopt the outside view and consider the opposite" [159]. For example, anchoring effect could be drastically reduced by asking people to consider arguments that are inconsistent with the anchor [189]. However, too many alternatives can also bring unnecessary burden to the decision-making – research shows that increasing the number of alternatives from 2 to 3 can greatly improve the quality of decisions, while when there are too many alternatives, the decision making quality deteriorates as much less time and effort are invested in evaluating each alternative [87]. Therefore, different strategies for evaluating the alternatives should be adopted for different contexts.

In decision making theory, there are two types of strategies to choose between alternatives. To make an optimal choice between a set of alternatives, ideally, we should be able to come up with an explicit set of criteria for the decision, and combine all the criteria together through some models – it can be as simple as *weighted additive* of the criteria [61], or more complex models such as *Analytic Hierarchy Process* [220]. Such style of decision making is known as *compensatory strategy*, which aims to evaluate the alternatives by combining all information and consider the trade-offs between different factors [192]. However, for decisions in VA, compensatory strategies can be hard to implement – the criteria for choosing which part of the data to explore first can be hard to determine, the number of alternatives for tweaking certain parameters for an algorithm can be infinite, and the value of putting the calculated results into certain type of visual encoding often can not be evaluated unless already visually presented. Moreover, as the decisions in VA are usually easily reversible, the "trial-and-error" type of interaction is commonly adopted in users [276] – in this case, analysts might temporarily settle with a "good enough" decision for the undertaken tasks, so the analytical process could move forward. Therefore, a *non-compensatory* strategy that does not consider all information but eliminates alternatives that do not meet some particular criteria is often used. For example, *"Take-the-Best"* is one of the most prominent heuristics which use the "best" piece of available information that discriminate two alternatives when analysts are making a binary choice [151]. In a similar spirit, *Elimination by Aspects (EBA)* considers the most important attribute among the two or more alternatives and eliminate the ones that do not meet certain cut-off value, then the next most important attribute is considered, until only one alternative left [151]. In the case where an excessive number of alternatives can not be avoided, the strategy of choosing the alternatives becomes increasingly important. Non-compensatory strategies like EBA can often help to filter out some potential candidates before going into deeper evaluations. For example, disjunctive rules accept alternatives that fulfill any requirements set on their attributes, conjunctive rules allow alternatives that fulfill all of the requirements, while lexicographic rules rank the importance of the attributes and "take-the-best" from the alternatives when one alternative is significantly better than the rest in any of the attributes [182] .

In real life, we often practice decision making in a hybrid manner – for example, when we buy on computer online, we might first filter the price range and certain specifications (non-compensatory strategy) to narrow down the candidates, and then evaluate the last few alternatives thoroughly by looking through all relevant product information and even comments or reviews (compensatory strategy). Decision support resembling non-compensatory strategies can also be seen in recent research for VA. For example, Tableau's "Show Me" [162] and suggests preferable visualizations based on the selected data to analyze, Draco [183] provides alternatives using users' specification as well as constraints from visualization design knowledge (see Figure. 8.4), and Voyager [280] recommends related views based on users' specified view (see Figure. 8.5). These research all utilize multiple views to exemplify the alternatives to choose from and use recommendations to help users avoid some flawed alternatives (non-compensatory strategies), therefore improve the trustworthiness of users' decisions.

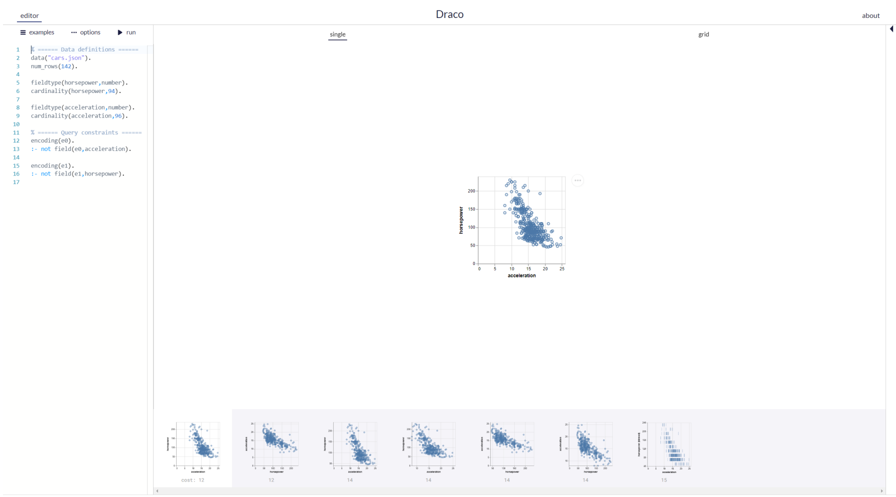


Figure 8.4: Draco [183] utilizes visualization design knowledge as a set of constraints, and recommends visualizations at the bottom of the currently specified view based on such constraints and the specifications from users to promote effective encoding.

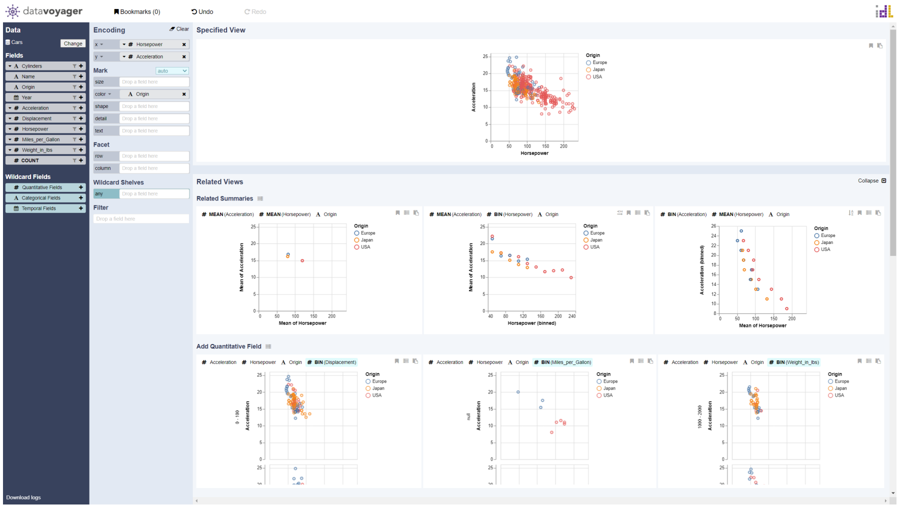


Figure 8.5: Voyager [280] provides related views at the bottom of the specified view that suggest relevant visualizations based on the current visual encoding selected by users.

In contrast, compensatory strategies examine all the possible variables and combine them in a structured way, therefore can provide reliable and stable pathways for making trustworthy decisions and greatly improve the comparability or reproducibility of the VA process. Implementing compensatory strategies in VA, however, remains a formidable challenge as to concretize the specific criteria regarding each decision in VA to consider and structurally present these criteria with regard to each alternative.

8.3 DECISION RATIONALITY AND DUAL PROCESS

The non-compensatory strategies mentioned in Sec. 8.2 that only consider a limited subset of information also reflects another important concept in decision making and economics – bounded rationality. It asserts that humans make inferences with limited time, knowledge, and resources, therefore look for alternatives that "satisfice" – satisfy and suffice rather than a globally optimal one [235]. Proponents for these "fast and frugal" heuristics, often non-compensatory strategies, such as "Take-the-Best" and EBA, argue that they are not necessarily irrational, and showed that they can outperform in both speed and accuracy in some instances through a series of experiments [89, 170]. However, issues would also arise when taking the non-compensatory strategies to the extreme – seeking only information that confirms one's assumptions (confirmation bias), leaning to certain options that they were more exposed to (mere-exposure effect) or recently exposed to (recency bias) [66], then the decisions could be extremely biased and potentially untrustworthy. Previous research in decision making also shows that even though experts are good at identifying the important attributes about the alternatives for accurate and trustworthy decision-making, they tend to be poor at combining and synthesizing these attributes [72]. This is where such cognitive biases and pitfalls come into play.

This extreme side of non-compensatory strategies and bounded rationality also relates to an important theory in decision making – Dual Process. *Dual Process* theory proposes that human reasoning consists of two relatively independent type of processes: type 1 – an fast, unconscious and implicit process with large capacity, and type 2 – a slow, conscious and explicit process limited by the capacity of working memory [79]. For example, to decide if a patient should go to a coronary care unit or regular bed, the doctor can use their past experiences (type 1) and/or medical instruments (type 2). Such dual process is also echoed in stereotype and prejudice studies – one of the most significant studies in the field by Devine concluded through a series of experiments that stereotypes can be unconsciously activated and applied (type 1) regardless of one's personal belief, while given enough mental resource and motivation, one with low prejudice level can inhibit the use of stereotype with their controlled cognitive process (type 2) [64].

In Visualization research, Padilla et al. proposed a cognitive framework in decision making with visualization based on dual process theory, and connected different thinking process in visualization with the two types of processing [200]. From the perspective of trust in VA, type 2 processing is more trustworthy, as the decision would be more structured and considered with more information. Such processing can be elicited with structured decision making strategies such as compensatory ones. However, type 1 processing still has its important value in efficiency, which is essential to ensure relatively good usability. Therefore, it becomes an important task for VA researchers to investigate when should which type of processing to be activated, and how to leverage type 1 processing to ensure interaction and decision efficiency while avoiding potential pitfalls.

8.4 DECISION ANALYSIS

Different from decision strategies, decision analysis aims to model and predict human decisions [192]. As Booth et al. pointed out in their paper on decision making modeling [20], relevant VA research, from Van Wijk's Value model, Green et al.'s Human Cognition Model, to Sacha et al.'s Knowledge Generation Model, has primarily focused on a *normative* approach – discussing what a rational human *should* logically do, in another words – the "best practices" in visual analytics. Particularly, Van Wijk's value model emphasizes that great visualizations lie in obtaining highly valuable knowledge with low cost of time and money [265]. From both theoretical and empirical analysis, other research in VA also attempt to formalize users' reasoning and sense-making process in terms of actions, tasks, and corresponding goals. This idea resonates with the prime example of a normative decision making model – expected utility hypothesis:

Expected Utility of each alternative is computed by the weighted sum of the utilities of its all possible outcomes, and it is assumed that rational individuals will maximize the expected utility and therefore choose the alternative with highest value of expected utility [15]. For example, when designing a visualization with the property of different cars from different country origins, an experienced user would most likely choose color or texture to encode the country origin property instead of size for a higher expressiveness, therefore higher expected utility. Although a normative approach does provide insights on the maximum potential of utilizing VA systems and what users should do to achieve that, researchers also become increasingly aware that what users *actually* do in reality is often based on heuristics and can deviate from the rational and logical course of reasoning. Perceptual differences, knowledge gaps, cognitive biases and situational factors could all contribute to such deviation [103].

To formalize such heuristics-based approach, research in decision making developed a different type of model – *descriptive* decision theories – to capture how people *actually* make decisions. Among them, *Prospect Theory* is the most prominent model for descriptive decision analysis – it maintains the idea of maximizing some form of expectation, but the expected utilities regarding the outcomes are considered relatively to a reference point (e.g., current wealth in the case of investing or betting) and cognitively distorted in a non-linear and asymmetric manner regarding gain and loss. Figure. 8.6 exemplifies the value of losing $100 is more significant than gaining $100. In situations with risks and uncertainties, human tend to be more risk-seeking when the choices lead to or are framed as losses, while more risk-averse when it comes to gains [132, 263]. Such dynamics with risk are important when people make decisions with the results of visualization – making life-and-death medical decisions, investing a huge amount of money, or developing policies that might influence the life of millions. Previous studies also show that high quality visualizations can well enhance the communication of risk, while perceptual errors can still arise and lead to the distortion of probability estimates [77, 81]. However, when analysts make each decision in VA, most of them are of low risk and easily reversible – one can always zoom out from a zoomed-in area of data, try out another algorithm, or use a different chart and visual encoding. This therefore can make VA decisions fundamentally different from many other decision-making scenarios – in many VA systems, users are

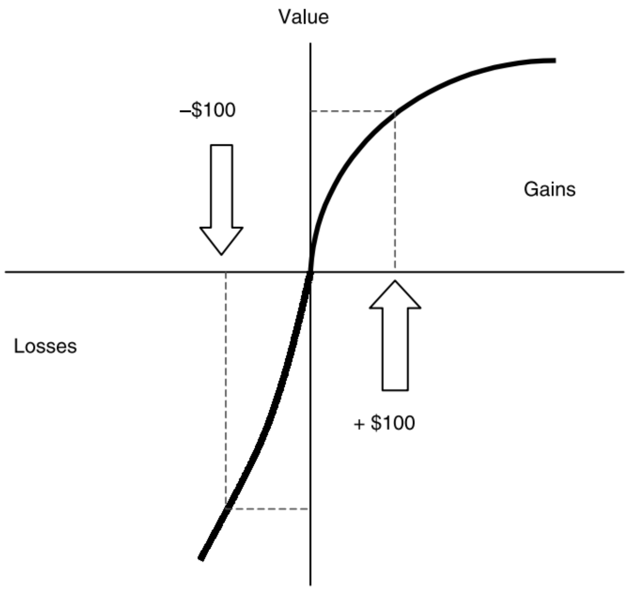


Figure 8.6: The value function of prospect theory [132, 192] where the value of loss is more significant than the same amount of gain.

often encouraged to explore and try out possible analytical paths – as the effort to recover from mistakes can be very low, and the risk of making a decision now is therefore nearly non-existent. Essentially, the effort going into making a decision is to reduce the risk of making an erroneous decision [53] – too much effort could be costly, while too little effort can greatly increase the risk. In the case of many potentially temporary and reversible VA decisions, investing too much effort is not worthwhile. This not only relates to the bounded rationality we discussed before – users often have limited time and resource to invest in making each visual analytical decision, we also need to consider with the "trial-and-error" style of decision making, how can we create feedback to users to help them make trustworthy decisions after the error.

With regard to risk, previous research also pointed out that the perceived risk and the actual risk of a decision can greatly differ from each other. Slovic et al. explain how risk is constructed in two ways – feelings as one's instinctive and intuitive reaction to danger, and analysis as one's logical and cognitive deliberation on risk management [239]. In particular, risk as feelings, or affect, can be mixed or influenced by other feelings, such as benefits – when a decision is framed as beneficial, the positive affective evaluation will lead to an inference of lower risk and decrease the perceived risk, and vice versa. Conversely, when an alternative is linked to negative affect, the corresponding perception of risk can increase and therefore overrated. With increased perceived risk, analysts can become more reluctant to make decisions and interact with the VA systems. This is also closely related to what our previous discussion regarding calibrating trust – the perceived risk of a decision also needs to be calibrated with regard to its actual risk for users to have the calibrated level of trust [103].

In addition to risk aversion in prospect theory, descriptive decision analysis also models many other issues regarding decision pitfalls, such as framing effect [71, 263], anchoring effect [262] and ambiguity aversion [107]. These predictive models can greatly contribute to warning flawed decisions made by users and highlight pathways for users to make trustworthy decisions.

8.5 CONCLUSIONS

In this paper, we advocate for a research focus on making and trusting decisions "in" besides "with" VA. To this end, we inspected relevant decision making theories – namely decision strategies, bounded rationality and dual process theory, as well as decision analysis models – with regard to making decisions in the VA process, and discussed their potential for making trustworthy decisions. From these discussions, we conclude the following potential research pathways for trustworthy decision making in VA:

First, both presenting a number of alternatives to choose from and providing relevant information regarding these alternatives contribute to trustworthy decision making. This not only helps users make more informed and trustworthy initial decisions, but also enables users to trace the provenance of their decision and analytical process, which is extremely important in an iterative VA process where users might later adjust their previous decisions. However, with bounded rationality, users tend to utilize their instinctive processing to capture limited amount of information. This also needs to be considered with regard to how to guide such processing towards more trustworthy decisions.

Second, descriptive decision analysis models can help to understand and highlight errors in user decisions. However, these models are yet to be adapted to the specific natures of making decisions in VA, for example – VA decisions are usually of low risk, easily reversible and iterative. Further inspections on these extended research from decision making community, such as on framing effect, anchoring effect, and ambiguity aversion [107, 262, 263], can greatly benefit VA research. Normative decision making model and different decision making strategies can be utilized to guide users to make more structured and trustworthy decisions. Normative decision model provides fundamental theories regarding expected utility that can be utilized in decision making strategies, and both compensatory and non-compensatory strategies also enable more accurate and trustworthy decision making with structured criteria or heuristics.

Finally, we can observe a common pattern of "intuition vs. logic" dichotomy from the decision making theories (see Figure. 8.1). However, both our discussion and research in decision making point out that decisions are usually not clean-cut through these diverging lines, and both sides of the models are very often combined together for most decisions. In addition, although these more intuitive models, strategies and processing can lead to some common pitfalls of cognitive biases, many decision making researchers also pointed out the high accuracy and efficiency of these intuitions are essential to human decision making. Therefore, it is vital for VA researchers to recognize the importance of facilitating and utilizing these intuitive approaches while avoiding the pitfalls they might bring along.

9

PAPER C: GUIDANCE METHOD

Designing and Providing Visual Analytics Guidance through Decision Support

Wenkai Han and Hans-Jörg Schulz

Department of Computer Science, Aarhus University, Denmark

ABSTRACT

Guidance in visual analytics aims to support users in accomplishing their analytical goals and generating insights. Different approaches for guidance are widely adopted in many tools and frameworks for various purposes – from helping to focus on relevant data subspaces to selecting suitable visualization techniques. With each of these different purposes come specific considerations on how to provide the needed guidance. In this paper, we propose a generic method for making these considerations by framing the guidance problem as a decision problem and applying decision making theory and models towards its solution. This method passes through three stages: (1) identifying decision points; (2) deriving and evaluating alternatives; (3) visualizing the resulting alternatives to support users in comparing them and making their choice. Our method is realized as a set of practical worksheets and illustrated by applying it to a use case of providing guidance among different clustering methods. Finally, we compare our method with existing guidance frameworks to relate and delineate the respective goals and contributions of each.

9.1 INTRODUCTION

Guidance in visual analytics (VA) has received increasing attention in recent years. Defined as *"a computer-assisted process that aims to actively resolve a knowledge gap encountered by users during an interactive VA session"*, [41] guidance aids users in producing analytic results, generating new insights, and eventually building new knowledge. [37]

Despite being a concept recently introduced to VA, the practice of using different forms of guidance to support users' analysis processes has been widely adopted in many VA systems – from data [259] and visualization exploration [152, 280] to model building [186] and reasoning. [255] The wide and diverse usage of guidance validates the usefulness of the concept and provides a solid foundation for research on guidance mechanisms. With a clear characterization and a broad range of applications, recent research extends guidance beyond resolving knowledge gaps, [51] discusses when guidance should be used, [38] and lists the considerations to be made when designing guidance. [37] However, a generic guidance method for designing and providing guidance across its various applications and goals is still missing.

Facing this challenge, the main contribution of this paper is a step-by-step process to derive the practical "how" of providing guidance. To this end, we reformulate the guidance problem as a decision making problem and apply decision support models towards its solution. Our method passes through the following three stages:

1. Stage 1 identifies the decision points in VA processes where guidance is needed regarding the data, algorithms, visualizations, and reasoning.

2. Stage 2 utilizes multiple criteria decision analysis (MCDA) [275] to evaluate the alternatives to choose from in these decision points in order to generate guidance.

3. Stage 3 uses composite visualizations of multiple alternatives for inspecting and comparing the resulting alternatives to guide the users in their decisions.

In addition to its application-agnostic coverage of process, goals, and conceptual levels, there are various benefits of utilizing this method to provide guidance. From the perspective of designing VA systems, our method allows for quickly realizing and testing if and how guidance might work for the system. Once realized, it can also be more easily adapted and re-used in other systems. Finally, any scoring or ranking metric can be directly included as a criterion in MCDA models, making it backwards compatible to existing guidance implementations as well as to a wide range of utility and quality measures for data, algorithms, and visualizations. From the perspective of using VA systems, our MCDA-based guidance exposes the criteria and their weights to the users, which makes the generated guidance more explainable. It further allows for adjusting these weights to user- and scenario-specific needs, making the provided guidance flexible and adaptive. Finally, framing and presenting guidance consistently through criteria and weights provides for a uniform guidance experience across domains and applications.

9.2 BACKGROUND AND MOTIVATION

To explain the rationale behind our guidance method, we present the research background and motivate the reasoning behind our method from three perspectives: guidance for resolving knowledge gaps, generic guidance beyond knowledge gaps, and guidance through decision support.

9.2.1 *Guidance for Resolving Knowledge Gaps*

Framed around the concept of knowledge gaps, guidance supports users in their VA processes to overcome hurdles and successfully proceed in their analyses. In the following, we present the research on VA guidance from this perspective and provide our motivation for a generic guidance method that goes beyond knowledge gaps.

9.2.1.1 *Background*

Formally characterized in VA by Ceneda et al., guidance is usually framed around resolving *knowledge gaps* encountered by users in VA processes. [41] These knowledge gaps can be of different types, either *target-unknown*, where the optimal solution of a VA problem is unclear (I-know-it-when-I-see-it), or *path-unknown*, where the sequence of actions to reach a known target is unclear. Guidance can also be of varying degrees, ranging among orienting, directing, and prescribing; and lie in different domains, from data and tasks to VA methods and knowledge management. [41] Further research has brought forth some considerations and tools for designing guidance. A decision tree was proposed for deciding if guidance is needed and to what degree. [38] Five key requirements for effective guidance were established – namely for guidance to be available, trustworthy, adaptive, controllable, and non-disruptive. [37]

The conceptual space covers a wide range of approaches that have proven useful in assisting users to resolve their knowledge gaps. [39] These approaches come in various guidance degrees and guide users in different domains. *Orienting* users towards regions in *view space*, Gladisch et al. provide visual cues pointing to where potentially interesting data points lie based on a degree-of-interest function. [92] *Orienting* users towards different levels of granularity in the *data*, Luboschik et al. provide heterogeneity-based guidance that indicates hidden details at higher levels of granularity and thus guides users to zoom-in for a closer inspection. [160] *Directing* users in carrying out their analytic *tasks*, Streit et al.'s Stack'n'Flip approach provides a guided view of the user's analytic workflow and offers subsequent analysis steps to be taken. [255] *Orienting* and *directing* users among *VA methods*, Müller et al.'s Morpheus guides users through multiple parameters to choose the best ones for subspace clustering. [186] *Directing* and in part *prescribing* user actions in the space of generated *knowledge and insights*, the Nugget Management System from Yan et al. suggests and refines valuable information (nuggets) based on user interest. [287]

9.2.1.2 *Motivation*

The outlined research provides high-level guidelines regarding guidance design principles as well as context-dependent examples of how to guide users in VA. However, for VA designers to effectively implement guidance, a concrete method to produce guidance from end to end – i.e., from the specification of the guidance problem through requirements all the way to generating and presenting guidance to the end user – that is independent of the guidance domain or scenario is still lacking. Specifically – How to systematically identify the points where guidance is needed? How to compute and generate guidance with a unified underlying mechanism? How to present and adapt guidance according to the context? These questions motivate our guidance method that can be used in various guidance scenarios.

9.2.2 *Generic Guidance beyond Knowledge Gaps*

The utility of guidance goes far beyond knowledge gaps. In the following, we present related research in guidance supporting this perspective and motivating a guidance design framework that caters to a variety of guidance aims including, but not limited to knowledge gaps.

9.2.2.1 *Background*

Collins et al. aptly observed that the goal of guidance can go beyond resolving knowledge gaps, including to inform, to mitigate bias, to reduce cognitive load, for training, for engagement, and to verify conclusions. [51]

Indeed, when VA experts have sufficient knowledge to conduct their analyses, guidance can nevertheless support them by providing important meta-information, keeping track of their analyses, and making suggestions to reduce their cognitive load and improve their efficiency. [97, 130, 280] Furthermore, no matter how knowledgeable a VA expert is, they might still be subject to various cognitive biases in their analyses – especially subconscious ones [66] – and guidance can also help to combat these biases through revealing them and providing suggestions when indicators for bias are detected. [270] Other research has also discussed guidance goals, such as attention management by guiding users to views that are currently important, [273] supporting user learning and training with onboarding guidance that walks users through visualization elements, [253] engaging users by prompting them when they are inactive, [52] as well as aiding the verification of analyses through monitoring of analytical hypotheses. [130]

9.2.2.2 *Motivation*

To construct a generic guidance method that is applicable to a wide range of guidance goals, we need to base it on a concept that also includes the additional goals of guidance beyond resolving knowledge gaps, such as to reduce cognitive load, to mitigate bias, and to verify conclusions. However, critical challenges lie in not only finding such concept and establishing its conceptual connections with guidance, but also integrating its existing theories as well as applications with guidance in order to practically build our guidance method.

9.2.3 *Guidance through Decision Support*

To yield the missing end-to-end guidance framework that is inclusive of the extended set of guidance goals, a new perspective on user guidance in VA is needed. This is where decision making theory in general and decision support in particular come into the picture. In the following, we present related work in decision support, and motivate our method for using decision support to design and provide guidance.

9.2.3.1 *Background*

A large body of decision support research has focused on a similar set of goals as the one that Collins et al. proposed for guidance. These include, for example, to provide important information relevant for the decisions, [269] to reduce errors and mitigate various biases, [88] and to alleviate mental workload when the decisions are complex. [212]

Research on Decision Support Systems (DSS) provides us with useful tools to realize a generic guidance method. The Handbook on Decision Support Systems provides a useful overview of this topic. [31] Its chapter on DSS architectures and types summarizes the four basic components of a DSS, including language (input), presentation (output), knowledge (database), and problem-processing (model) (sub-)systems. Relating these components to the goals of guidance, the knowledge components contain important information to inform users and store expertise that reduces their cognitive load; [156] the problem-processing components structure and model users' decision making processes, helping users to combat their biases, and ease their cognitive effort; [88] the language and presentation components also aid users' mental work, while promoting users to be engaged in the decision making processes through interactions. [153]

Among these four components, those focusing on problem-processing are particularly relevant to guidance generation, as they provide a mechanism to analyze decisions and generate evaluations of alternatives, which is essentially what guidance does. Multiple criteria decision analysis (MCDA) is a commonly used approach for problem-processing. [275] It takes multiple quantifiable metrics as evaluation criteria and evaluates alternative decisions based on these criteria. As the underlying criteria can be easily exposed to and manipulated by users, MCDA methods allow for communicating and potentially even changing the mechanism providing the guidance.

9.2.3.2 *Motivation*

The similar sets of goals of DSSs and guidance make DSSs a suitable method to generate guidance in support of any decision making problem – be it which data to look at, which algorithm to choose, or in which direction to pan. Particularly MCDA methods are promising in this regard. However, the critical challenge still lies in how exactly to map the decision support research to specific steps for constructing a generic yet practical guidance method. This is where our work contributes by connecting research in decision support and VA to build a generic method for guidance design with concrete steps to follow.

9.3 OVERVIEW OF THE METHOD

This section gives an overview of our guidance method through decision support. We first conceptualize guidance as a decision making problem, before structuring our method based on the decision making process.

9.3.1 *Guidance as Supporting Decision Points*

When observing situations in VA in which guidance is needed, it is noticeable that users are often faced with making decisions among multiple alternatives: Which data (sub)set to use? What algorithms and parameters to choose? What visual encoding to use? Where to start or proceed with an analysis? – These decisions are inherent in VA: if no human decisions were required, then the analyses could be fully automated and neither a human user in the loop, nor guidance would be needed. Therefore, we can see the existence of "decision points" as a prerequisite for needing guidance. Knowledge gaps can then be seen as a common issue, among many others, that might arise at these decision points – when users lack the knowledge to make their decisions in the analysis, guidance can be used to provide such knowledge.

This perspective of an analysis workflow as a series of decisions among multiple analytic alternatives is echoed in recent research by Liu et al. [155, 157] They studied how researchers experiment with different paths when analyzing data and identified the points where these alternative paths fork as "decision points". Their work provides a fundamental understanding of what constitutes decision points in analytical work and how analysts reason in their decisions. However, how to support these decision with guidance is an open question. In addressing this question by connecting decision support systems with guidance, we provide a novel and tangible path for guiding users.

The concept of decision points also covers the realm beyond knowledge gaps. At decision points, the users' preconceptions might lead them to choose certain analytical paths to confirm their hypotheses (confirmation bias), where guidance can suggest alternative solutions and mitigate such biases; or particular analytical decisions might be cognitively complex and demanding, where guidance can facilitate them among the space of alternatives and reduce the cognitive load. Therefore, decision points are a fitting concept to capture situations for which guidance is needed and can be provided through decision support.

In short, we extend the guidance concept to decision points and re-frame the guidance problem as **providing users with decision support when they are faced with decision points in VA that involve multiple alternatives**.

9.3.2 *Structure of the Guidance Method*

Our method for designing and providing guidance is based on this re-framing into a decision making / decision support problem, whose overall structure is outlined in this section.

We look at decision points in VA from each of the three stages of decision making processes proposed by Herbert Simon – `intelligence`, `design`, and `choice`. [236] Simon's model has been widely studied, including in the context of visualization tasks. [67] The `intelligence` stage refers to recognizing the conditions calling for decisions, the `design` stage refers to the development and evaluation of the alternatives, and the `choice` stage refers to choosing the desired alternative(s) based on the evaluation results. These three stages are visually presented in Figure 9.1.

The focus of our method in the `intelligence` stage is to *detect and assess the decision points where guidance is needed*. This stage starts with

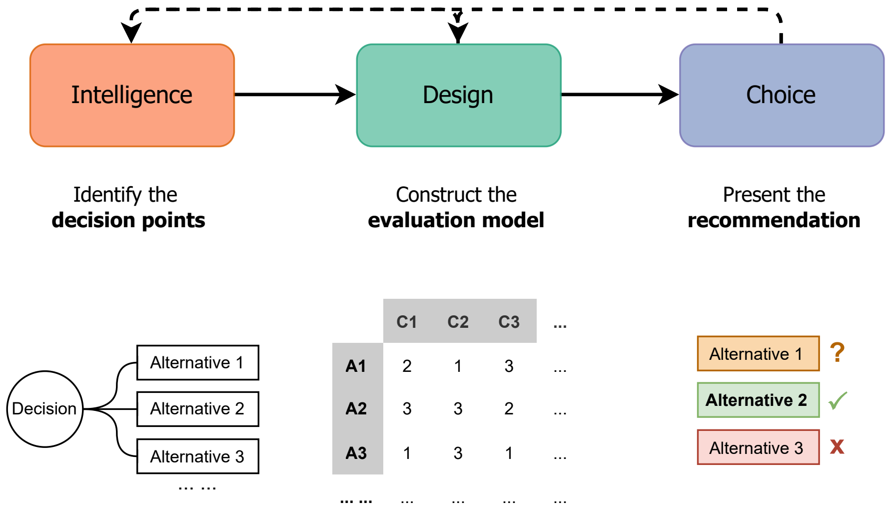


Figure 9.1: The three stages in decision making process by Simon [236] with illustrated explanations of the three stages.

an inspection of the context in which the VA system is being used. This helps VA designers to systematically understand the conditions under which guidance might be needed. With the context recognized, we can then systematically identify the decision points calling for guidance. Finally, to evaluate if guidance is actually needed and prioritize these identified decision points, we assess the need for guidance at each decision point.

In the `design` stage, the focus is on constructing the mechanism for guidance generation through *developing and evaluating the alternatives* for each decision point that calls for guidance. Continuing with the assessment of decision points produced in the `intelligence` stage, we first aim to recognize the space of alternatives for each decision point by specifying the number of alternatives along with a list of examples to be evaluated. This allows VA designers to more tangibly consider how to evaluate and present these alternatives in the later steps. Thereafter, we discuss how evaluation criteria can be produced for the alternatives. Finally, to evaluate them, we introduce MCDA and how it can be applied and adapted to varying guidance degrees.

The main challenge at the `choice` stage of our guidance method is to *present and adapt the guidance output* for each decision point. The first step in this stage carries over the results from the `design` stage by recognizing what data to present in the guidance output. Combining this data with the corresponding guidance degree and number of alternatives, we then consider how to compose the presentation of alternatives in order to produce a guidance output suitable for the given guidance scenario. Finally, to allow the guidance to be adaptive and flexible, we inspect how to adjust the produced guidance in different contexts.

With the three stages described above, we have outlined our generic guidance method framed around decision points on an abstract level. The conceptual connections between the key concepts in the initial characterization of VA guidance and our method in are illustrated in Figure 9.2. In the next three sections, we further detail the concrete steps in each of these three stages.

It is worth noting that the following design process formulates a middle ground that we expect to be applicable to most, but not necessarily all guidance scenarios. The reason is that guidance can vary greatly in its complexity depending on its context. In a simple analytical process with well-established "best practices", guidance can be provided through a manually authored workflow with alternative paths without designing an elaborate evaluation model. Whereas in complex analysis settings where the alternatives' evaluation criteria are challenging to define or compute, our design process may have to be reiterated multiple times and the evaluation model may include more complex considerations than can be expressed by weights alone. Therefore, the following description should not be used dogmatically as a fixed end point of all design considerations, but flexibly as a starting point from which to tailor a sensible design process for a guidance problem at hand.

9.4 stage 1: intelligence – decision points calling for guidance

From the perspective of guidance as decision support, it is vital to first recognize the decision points for which to provide guidance.

In the following, we illustrate the steps in this `intelligence` stage by first articulating the context of use, then identifying the decision points in such context, and finally assessing the need for support in each of the decision points in order to prioritize them when implementing guidance. These three steps are listed in Figure 9.3.

9.4.1 *Step 1.1 – Analyze the context of use*

9.4.1.1 *What*

The context of use captures the conditions under which a product is used. Analyzing the context of use of a VA tool provides important information for devising effective guidance in later steps. For example, the offered guidance may differ depending on whether a VA tool is being used within or outside of its intended context of use.

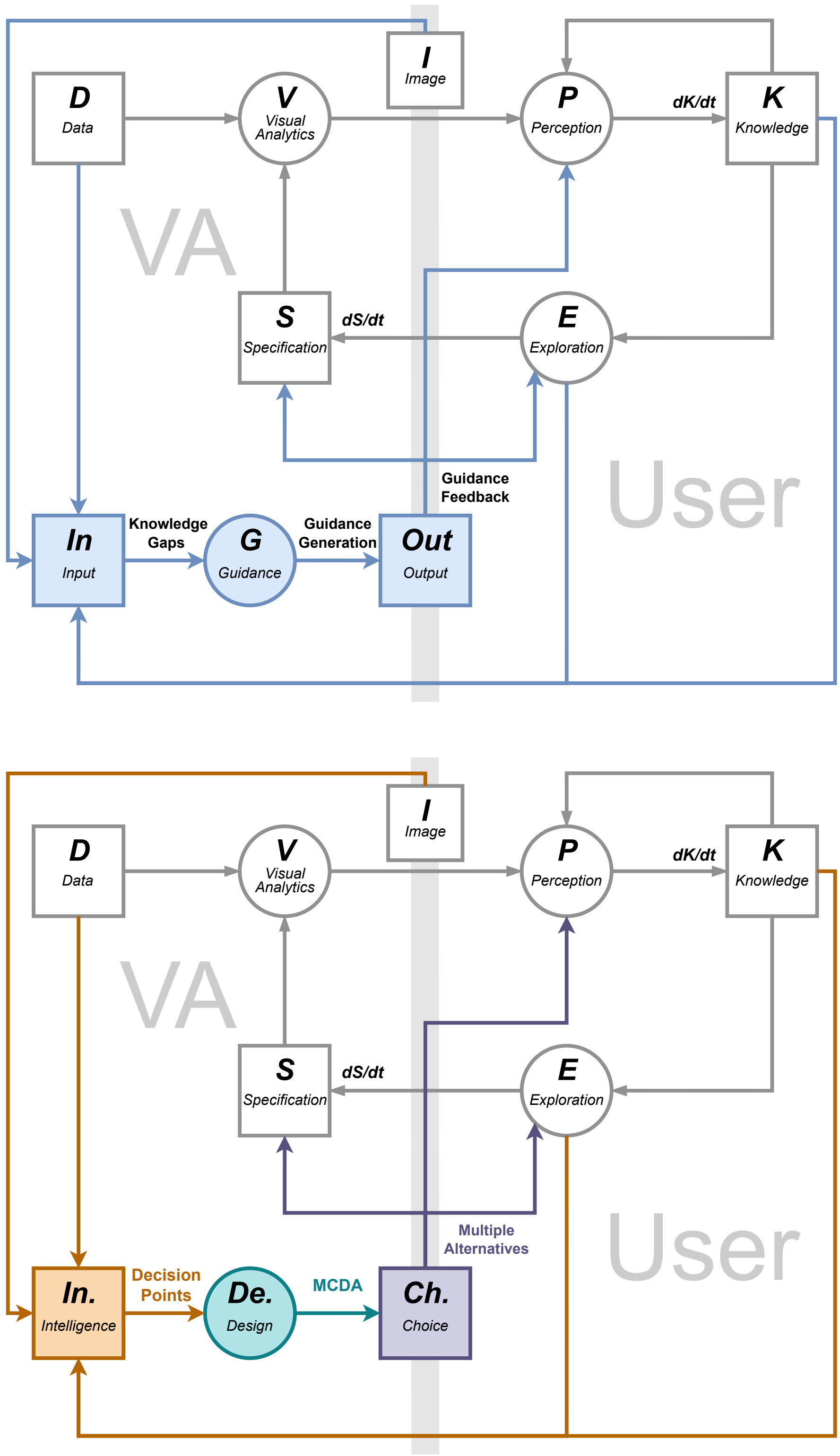


Figure 9.2: Adapting the guidance model by Ceneda et al. [39, 41] (top) to reflect the stages of decision making processes used in our method (bottom). For clarity, we subsumed the components of "history" and "domain" together with "data" and "knowledge" as "input", and the three guidance degrees as "output".

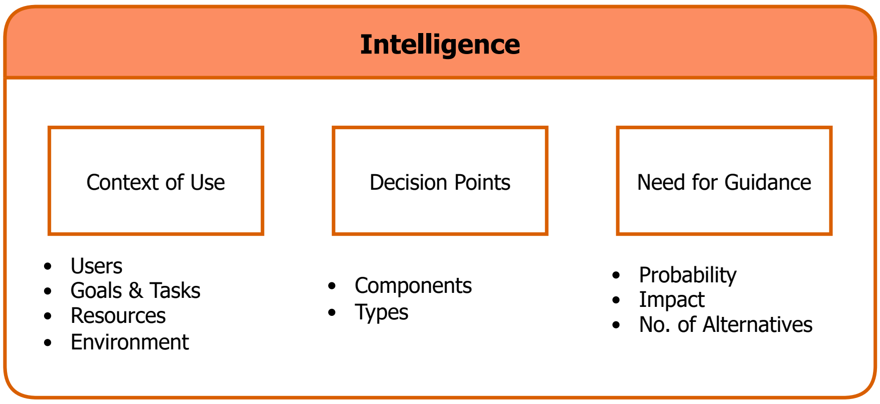


Figure 9.3: The steps in the `Intelligence` stage and the corresponding factors considered in each step.

9.4.1.2 *Why*

Context of use analysis is a prerequisite for designers to understand when, where, how, and by whom a system is being used to provide good usability. [249] This is also true for VA systems for which the context of use gives rise to the "conditions calling for decisions" [236] – i.e., the concrete situations in which guidance is needed. A thorough context of use analysis helps VA designers to more concretely identify the decision point (Step 1.2) and assess them (Step 1.3), as well as to build adaptive guidance generation (Stage 2) and presentation (Stage 3) grounded in knowledge about users, goals & tasks, resources, and environment.

9.4.1.3 *How*

Context of use is an important concept in HCI that can be interpreted from various perspectives. [69] Hence, the following considerations taken from Common Industry Format (CIF) for context of use descriptions (ISO/IEC 25063:2014) provide a least common denominator as a starting point for analyzing contexts of use, [248] but they should by all means be extended by additional, possibly domain-dependent considerations if these help to further pinpoint usage scenarios in which guidance may be needed.

- **Users** are persons directly interacting with a VA system. In particular, their level of expertise should be examined as users often need support when they lack knowledge or experience. [37]
- **Goals & Tasks** relate to the motivation and execution of visual analyses. For example, different goals – explore vs. confirm [229] – relate very much to different degrees of freedom in an interactive analysis and have thus implications for the degree of guidance needed – orienting vs. directing.

- **Resources** capture the boundaries of the analyses to be run. They include technical limitations such as computing power and available visualization methods, as well as virtual limitations such as cognitive strain and time constraints. Resources delimit the space of viable alternatives among which to guide users.
- **Environment** is a multi-faceted factor that includes technical, physical, social, cultural and organizational environments. For guidance, characterizing the technical environment is particularly important – e.g., describing how much visual support is already provided by a user interface.

Notably, context of use is often dynamic and subject to change according to the purpose of the system and the progress of development. Especially at an early stage of the system development, designers might not be able to articulate all the factors. Thus, the context of use analysis should be an iterative process and adapt to different usages.

9.4.2 *Step 1.2 – Identify the Decision Points*

9.4.2.1 *What*

Having established the context of use, we now identify the decision points within this context where multiple alternatives exist among which a user must choose.

9.4.2.2 *Why*

Decision points are essential for generating guidance (Stage 2) and presenting guidance (Stage 3), as these indicate where guidance may be needed. This step generates an overview of the various decisions users are facing in a VA tool without yet prioritizing among them, to ensure that decision points are not overlooked.

9.4.2.3 *How*

In the early stages of developing a VA solution, designers might not have direct access to users yet. In this case, decision points can be identified through a cognitive walk-through. At later design stages, decision points can be identified by involving users through interviews and contextual inquiries, possibly following structured protocols to assess challenging situations. [50] In both cases, designers need to closely inspect each component in a VA tool in order to comprehensively recognize the decision points.

To do so in a structured way, we propose the use of suitable task taxonomies for VA. This makes sense, as fundamentally any task carried out by the user involves a decision – e.g., Should I rather filter or sample the data to reduce it? Should I zoom-in here or there to see interesting details? If there was no decision to be made, the user would not have to carry out the task as the system could proceed by itself.

While many taxonomies exist for visualization tasks, [137] the literature is more sparse on VA tasks. After considering the taxonomies by Gotz and Zhou [97] and Heer and Shneiderman, [109] we settled on von Landesberger et al.'s taxonomy of VA interactions for its universality, high level of abstraction, and inclusion of analytic reasoning. [147] Slightly adapted to the context of decision making in VA, it breaks down into **components** and **types**. **Components** are the aspects of a VA system to which a task relates:

- Tasks relating to the **data component** deal with decisions on which data to use (e.g., subset selection, filtering) and how to use them (e.g., cleaning, transformation).
- Tasks relating to the **algorithm component** deal with decisions on how to process the data (e.g., which clustering algorithm) and how to parameterize the processing (e.g., distance metric, similarity threshold).
- Tasks relating to the **visualization component** deal with decisions on which visualization techniques to use and how to parametrize them (e.g., color mapping and axis scaling).
- Tasks relating to the **reasoning component** deal with decisions on which line of analytic reasoning to follow to yield insights (e.g., deductive reasoning to "detect the expected" or inductive reasoning for "discovering the unexpected") and how to carry it out. [257]

Types delineate between the fundamental *What to do?* and the subsequent *How to do it?* of a task. Concretely, these types are:

- Decisions on the **scheme** of a task – e.g., What data subset to analyze? What algorithm to choose? What chart type to use? These are fundamental decisions for or against principal options.
- Decisions on the **parameters** of a task – e.g., How to derive that subset? How to parametrize that algorithm? How to apply and fine-tune the chosen chart type? These are secondary decisions that follow from an already chosen scheme and that are needed to concretize and carry out that first decision.

9.4.3 *Step 1.3 – Assess the Need for Guidance*

9.4.3.1 *What*

To help VA designers to identify decision points that would benefit from guidance and prioritize them accordingly in the development process, we outline an assessment of the need for guidance among the decision points.

9.4.3.2 *Why*

Existing research on guidance emphasizes providing the right guidance at the right time and making sure the guidance is non-intrusive. [37, 51] Hence, it is important to assess the need for support in each of the decision points to avoid providing unnecessary guidance that may distract from the analysis or even disrupt the analysis flow instead of enabling it. Furthermore, there can be a large number of possible decision points in a VA system, and to provide guidance for all of them can be an arduous task. Therefore, the decision points need to be prioritized to create clear priorities for the guidance generation (Stage 2) and presentation (Stage 3).

9.4.3.3 *How*

The need for guidance at the decision points is influenced by many contextual factors. Hence, such assessment is ideally done together with the end users to ensure that it reflects the real-life experience of users through qualitative methods such as interviews or workshops as well as quantitative ones like surveys or user performance. Previous studies in guidance also indicate the potentials of usage logs from user interactions for inferring the need for guidance. [51] Without direct access to user information or logs, a cognitive walk-through or an internal expert review can be conducted to assess the need for support. [165, 261]

We formalize this process through an adapted version of risk assessment, which identifies and assesses potential risks at each decision point by quantifying the probability and impact of getting the corresponding analytic decisions "wrong". [148] Seeing guidance as a support mechanism to mitigate the potential risks at each decision point, we utilize a risk assessment scheme to evaluate the need for guidance.

- The **probability of a "wrong" analytic decision** is often characterized by *the lack of knowledge*, which can be decided by the various factors in the context of use. Previous guidance research has also discussed how to identify such knowledge gaps. [37] Other factors, such as the likelihood of cognitive biases may also factor into this probability. [74]

- The **impact of getting it "wrong"** is likewise a multi-faceted consideration based on how much the course and overall outcome of the remaining analysis workflow depends on this decision and how consequential a wrong result would be for the domain decision based on it – e.g., a wrong treatment decision for a patient would be more dire than a wrong ad-buying decision for a marketing campaign.
- Evaluating the **number of possible alternatives** is another important factor. If there are only a handful of possible options for a decision points, modern UIs with Undo/Redo functionality allow to quickly try them out before deciding for one without the need for an elaborate guidance scheme. Yet if there are many possible options, this is no longer viable, increasing the need for guidance with the number of alternatives.

After analyzing these factors, they can then be combined to produce the decision points inventory with a priority ranking. In risk assessment, the factors are usually rated on a quantitative scale, multiplied together, and combined in the form of a risk assessment matrix or a risk inventory. [148]

9.5 STAGE 2. DESIGN – MCDA TO GENERATE GUIDANCE

After identifying the decision points and assessing their respective need for guidance, we now specify the underlying mechanism that generates guidance for a decision point. To achieve this, we draw from the domain of decision support and propose to generate guidance through multiple criteria decision analysis (MCDA). MCDA integrates different criteria to evaluate the alternatives of decisions. [13] MCDA is a useful method for our goal of a generic mechanism to generate guidance, as any algorithm (e.g., heterogeneity-based guidance) or metric (e.g., degree-of-interest functions) can be easily incorporated in an MCDA model as criteria. [16]

For this stage, VA designers need to go over each decision point in the previously produced inventory according to the priority ranking. Additionally, as some of the decision points might be interconnected – for example, choosing a clustering algorithm and choosing its settings are often jointly decided, as they form scheme and parameters of the same decision – the guidance generation of such interconnected decision points can also be developed in conjunction.

In the following, we first recognize the space of alternatives that guidance generation should consider, then describe how the corresponding evaluation criteria can be produced, and finally discuss how an MCDA-based evaluation model can be built using these criteria to generate guidance. These three steps are listed in Figure 9.4.

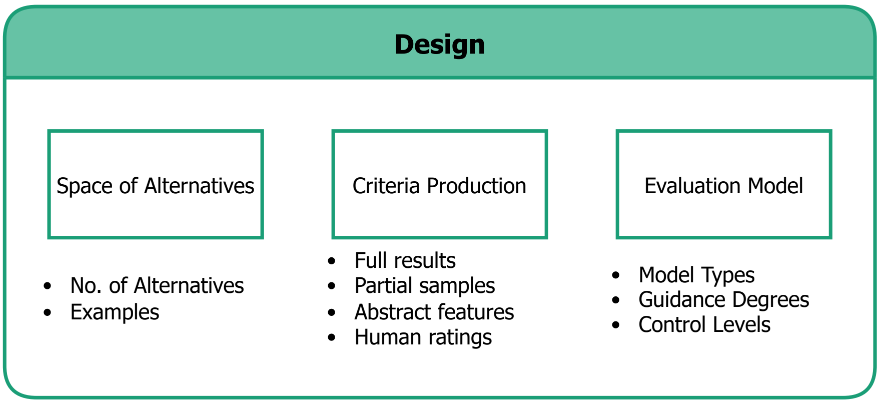


Figure 9.4: The steps in the `Design` stage and the corresponding factors considered in each step.

9.5.1 *Step 2.1 – Recognize the Space of Alternatives*

9.5.1.1 *What*

Although there might be a great number of available alternatives at each decision point, many of them may not be useful or possible to be considered. This step aims to recognize the space of alternatives that later feeds into the MCDA model in order to generate guidance.

9.5.1.2 *Why*

Recognizing the space of alternatives to be considered by guidance generation is essential for building the underlying mechanism that generates guidance. Specifically, the number of alternatives can influence how guidance should be generated and later presented in Stage 3. Moreover, recognizing some examples of the alternatives also helps VA designers to consider them in a more concrete manner and more easily identify the criteria to evaluate the alternatives.

9.5.1.3 *How*

To this end, we consider two elements that VA designers should inspect: the estimated number of alternatives and examples of alternatives. These two elements are considered under the constraints imposed by the identified context of use – users' goals and tasks that govern how open they are to explore different alternatives; available resources such as the set of implemented algorithms or the time available for inspecting different alternatives before having to make a decision; and the environment such as the available user interfaces that influence how open-ended the exploration of alternatives is. However, this step differs from the previous ones as we start to consider which of all the possible alternatives should feed into the MCDA model and be evaluated. Not all possible alternatives might be applicable or useful in a given context of use. And the subset of all those that are applicable may not be feasible to be evaluated due to time constraints or other limitations. And all those that are feasible to be evaluated may still be too many to then interactively inspect. To yield a clearer understanding of the practically relevant subspaces of alternatives, we look at the following:

- **The number of valid/useful alternatives** for each decision point, as their number often differs from the number of all possible alternatives from Step 1.3. Particularly, previous research in decision making shows that a higher number of alternatives can significantly improve the decision making quality, [201] while too many alternatives can also decrease decision efficiency and even lead to decision paralysis. [117]
- **Examples of the alternatives** illustrate what the alternatives look like for each decision point, such as names of different algorithms, different parameter range, or different encodings to be used for the visualization. These examples act as a concrete thinking tool for VA designers to consider potential alternatives under the contexts of use and later distill criteria that compare and evaluate them.

9.5.2 *Step 2.2 – Produce the Criteria*

9.5.2.1 *What*

The aim of this step is to produce the criteria to evaluate the alternatives for each decision point. These criteria will then later feed into an MCDA model and help to rank the alternatives and generate guidance.

9.5.2.2 *Why*

To evaluate the alternatives for a decision point, a way to judge them is needed. As we adopt MCDA models to calculate the overall evaluation from a set of input measures in Step 2.3, we need some form of quantifiable metric or quality measure to do so. It is also consequential for the subsequent presentation of guidance how these criteria are produced, as different methods to generate evaluation metrics may incur varying degrees of uncertainty.

9.5.2.3 *How*

To yield suitable evaluation criteria, we consider three kinds of measures based on the context of use and the number and types of the alternatives for the decision points.

- Measures **based on the full results** can be used when they are not too time-consuming to precompute. A wide range of such measures are available for representative data selection, [43, 60, 86] evaluations of machine learning, [125, 288] and quality metrics for data visualization. [12]
- Previews **based on partial results** for each alternative can be used to adapt to more time-sensitive contexts. Techniques like Progressive Visual Analytics can be helpful in such context of producing an early partial result and refining it over time. [4]
- If the methods above are not applicable due to the limitations in the context of use, predictive metrics **based on abstract features** of the alternatives can also be used, including data coverage, [270] algorithm runtime predictions, [119] and structure-oriented measures for visualizations. [18] This way, no precomputation of the result is necessary.
- Finally, when the metrics cannot be quantitatively and automatically produced or when they are too uncertain and imprecise, **human-rated** criteria and rankings can be used instead of computed measures. [22] Such human-rated criteria can be generated either through expert-rating or literature review.

Suitable criteria are chosen based on the identified contexts of the analysis (e.g., available time to generate and evaluate alternatives) and the decision point in question (e.g., algorithmic decision vs. visualization decision). To not only communicate the alternatives and their computed "goodness", but also their trustworthiness, their uncertainty may also be established and shown as meta-data for each alternative's rating. [19, 100] This is particularly important for criteria derived from partial or predicted results.

9.5.3 *Step 2.3 – Construct the Evaluation Model*

9.5.3.1 *What*

In this step, we combine the produced criteria for each decision point into an MCDA model to generate guidance.

9.5.3.2 *Why*

The evaluation model is an important element for generating guidance, as it forms the mechanism that produces the ranking among the alternatives that later enables guidance presentation (Stage 3). Here we consider how to build different evaluation models according to the varying degrees of guidance and user control, which allows the generated guidance to be adaptive and controllable.

9.5.3.3 *How*

MCDA methods come in various forms – Watróbski et al. summarized 56 different MCDA methods and discussed how to choose the corresponding method to support different decisions. [275] Overall, there are *three types of models* when it comes to combining criteria in MCDA:

- **Functional approaches** synthesize *quantitative criteria* into a single metric with assigned *weights* and optional *value/utility functions* for each criterion. [279]
- **Outranking methods** choose, rank, or sort the alternatives through *comparisons* between them based on a set of *quantitative or qualitative criteria* and corresponding *weights*. [22]
- **Decision rules** evaluate alternatives based on certain conditions and logic constraints that are often formulated as an *"if…, then…"* structure. [224]

To use MCDA models for generating guidance, the construction of them heavily depends on how the guidance should be used. Here we present two factors to consider – degree of guidance and level of user control.

MCDA models can adapt to varying *degrees of guidance*:

- As **functional approaches** produce the evaluation of alternatives as a single metric, they can be used to filter the alternatives with certain thresholds to provide orienting guidance, rank the alternatives by the produced metric to provide directing guidance, or select the highest ranked alternative to provide prescribing guidance. For example, the feature subset selection by May et al. filters and prioritizes features based on statistical ranking measures. [171]
- For generating guidance, **outranking methods** can be used similarly to functional approaches, except for their possibility of directly taking in qualitative criteria. Depending on the specific model that either ranks or discards alternatives, they can also either produce a ranking among the alternatives to provide directing guidance with the ranking result and prescribing guidance with the highest ranked alternative, or evaluate if the alternatives are acceptable to provide orienting guidance by filtering unacceptable alternatives.
- **Decision rules** often have the form "if alternative a is between x and y in criteria c, then a is a good enough alternative". Hence they can be used to filter alternatives for orienting guidance. For example, the underlying mechanism of "Show Me" in Tableau filters out visualizations not applicable to selected data based

on a set of similar rules. [162] Decision rules can also be used to direct users along branching analysis workflows as directing or prescribing guidance, like the Stack'n'Flip approach does. [255]

Guidance can further be controlled explicitly or implicitly through user input. [51] Here we discuss how MCDA models can be constructed to allow different *levels of user control*, including presets with no user control, inferences with implicit control, and direct input with explicit control.

- To construct a basic MCDA model, the appropriate weights, value/utility functions, and decisions rules for the criteria can be **preset** by the VA designers without user input, especially when users do not have detailed knowledge about these criteria and how to weight or constrain them.
- Furthermore, MCDA methods can also be controlled through **implicit inference** from user interactions. This is done through inferring some of the elements in the MCDA models based on user preferences produced by certain interaction patterns, such as mouse movement and user-generated materials. [27, 51] In functional approaches, value/utility functions can be elicited from a partial ranking of alternatives, [237] and weights can be elicited from the users' evaluation on the importance of the criteria. [211] Decision rules can also be flexibly modified, for example by inferring additional rules such as "if condition c occurs, then alternative a is more preferable than alternative b" from user choices made in the past.
- Experienced users who have abundant knowledge about these criteria and alternatives can also be exposed to the underlying evaluation mechanism and afforded with **explicit and direct control** of the elements in MCDA models. For functional and outranking approaches, this includes the criteria, their corresponding value entries, their weights, and optionally the value/utility function. For decision rules, this includes access to the rule set. Explicit control is also an important function for debugging guidance that does not work as expected.

9.6 STAGE 3: CHOICE – MULTIPLE ALTERNATIVE VIEWS FOR GUIDANCE OUTPUT

After the alternatives have been evaluated, users need to closely inspect and compare them in the context of their own domain knowledge and make their choices. To this end, the generated guidance must be communicated to the users for them to interact with and provide feedback to.

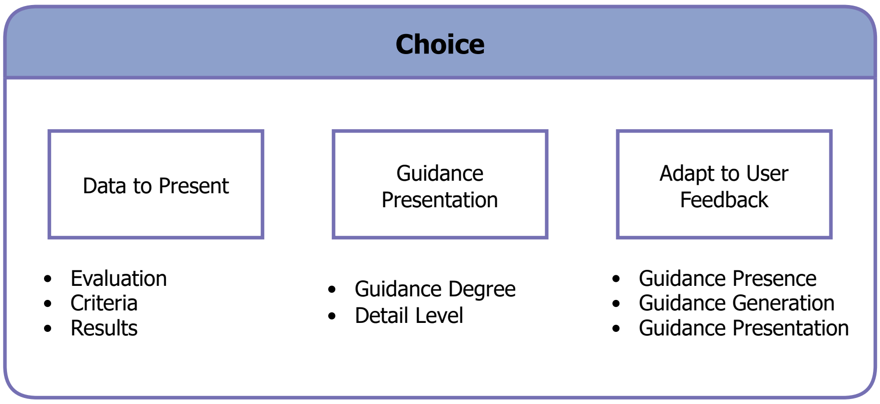


Figure 9.5: The steps in the `Choice` stage and the corresponding factors considered in each step.

In the following, we summarize the considerations in presenting guidance in ways that enable users to visually inspect, compare, and reason with these alternatives. While research has indicated a potential for providing guidance through other modalities than visual output (e.g., using vibrotactile feedback [104]) these approaches are still experimental at this point and require special hardware. Therefore, we focus on the visual channel to communicate guidance in this work, as it currently stands as the main modality for guidance in VA. We first recognize what data regarding the alternatives to visualize. We then discuss how to compose the guidance presentation based on the previous considerations. Finally, we summarize how the provided guidance can be adapted to user interaction and feedback. These three steps are listed in Figure 9.5.

9.6.1 *Step 3.1 – Recognize the Data to Present*

9.6.1.1 *What*

This step aims to specify what information/data about each alternative is relevant for users' choices and should thus be presented.

9.6.1.2 *Why*

What data to present depends on the available results produced at Stage 2 and decides the content and level of detail with which to present each alternative in Step 3.2. Hence, this step connects the guidance generation stage with the following step of guidance presentation.

9.6.1.3 *How*

In the previous `design` stage, we have generated and evaluated the alternatives through MCDA. This process produces different types of data:

- The main output of the `design` stage is the **evaluation** of the alternatives, which is the essential data element for presenting guidance. For orienting guidance, this is a list of all acceptable/-valid alternatives for the concrete analysis decision at hand. For directing guidance, this is a subset of the top-k best alternatives among the acceptable ones given as a ranked list to indicate the priorities among them. For prescribing guidance, this is the highest ranked alternative from that top-k list.
- The different **criteria** on which the evaluation model was based can also be shown. This can help users in understanding the characteristics of each alternative and uncover how the evaluation was generated. Yet, VA designers will need to consider if the users have the relevant background to interpret these criteria to avoid confusion or information overload.
- Some of these evaluation criteria might have been produced from **full or partial results** precomputed for each alternative during Step 2.2. Presenting these results can help users to inspect the detailed differences between these alternatives. For example, visualizing the resulting subsets after data selections can help users to gain an overview of the differences among these selections. Techniques known from comparative visual analysis, such as algorithmically-enhanced visual comparison, can further aid this inspection by deriving difference metrics. [146]

9.6.2 *Step 3.2 – Compose the Presentation of Alternatives*

9.6.2.1 *What*

Having described which data to present for each alternative, we then need to visually compose that data in order to display the guidance. In this step, we outline how to compose views of multiple alternatives based on the considerations we have made in the previous steps.

9.6.2.2 *Why*

Visualizing for guidance differs from other visualization in its fundamental characteristics of presenting and aiding the decision among multiple alternatives, such as the different levels of details and varying guidance degrees. Therefore, we outline the considerations for composing such visualizations that adapt to various characteristics of the guidance scenario at hand.

9.6.2.3 *How*

To visually present and compare the alternatives for VA decisions, multiple presentations of these alternatives need to be composed in a unified view. Techniques that combine multiple visualizations have been widely studied and given different names, such as coordinated and multiple views, [213, 215] composite visualizations, [126] and visual comparison techniques. [93] However, to visualize for guidance, different data and numbers of alternatives need to be presented and varying guidance degrees call for dissimilar presentations. Therefore, additional considerations need to be made in terms of presenting multiple alternatives.

To guide users through multiple alternatives, the *signification of guidance* can help to convey the relationships between alternatives and shift the focus to the important ones. For example, in Voyager, a "specified view" is put on top of other "related views", signifying a higher relevance of the "specified view" to the context of use. [280] Here we discuss the signification of guidance through the three guidance degrees – orienting, directing, and prescribing (see Figure 9.6).

- For **orienting guidance**, the alternatives should be visualized with the same visual importance and avoid implying any preference. However, there might still exist some underlying relationships between the alternatives that can help users to orient among them. For example, when the users are branching out into different paths of analysis, the logical and chronological relationships between alternatives can be indicated through new alternatives branching out from previous ones. [158, 274]
- For **directing guidance**, the key consideration is the underlying preference among the alternatives. Such guidance can be encoded in the order of which the alternatives are ranked, especially when they are presented as a list, where the linear order already implies a ranking of the elements – whether this is intended or not (cf. *position bias*). [282] To further emphasize the ranking, color, textual information, size, and/or animation can be used to indicate the preferred alternatives and the ranking among them. [51]
- For **prescribing guidance**, users are guided through a process where they can only accept the suggested alternative and navigate back-and-forth between different steps. However, additional alternatives can still be useful to show so that users better understand the context in which the prescribed alternative is generated. In this case, the additional alternatives can be presented in a de-emphasized way without interactivity to indicate that they are not available to be chosen and only shown as contextual information.

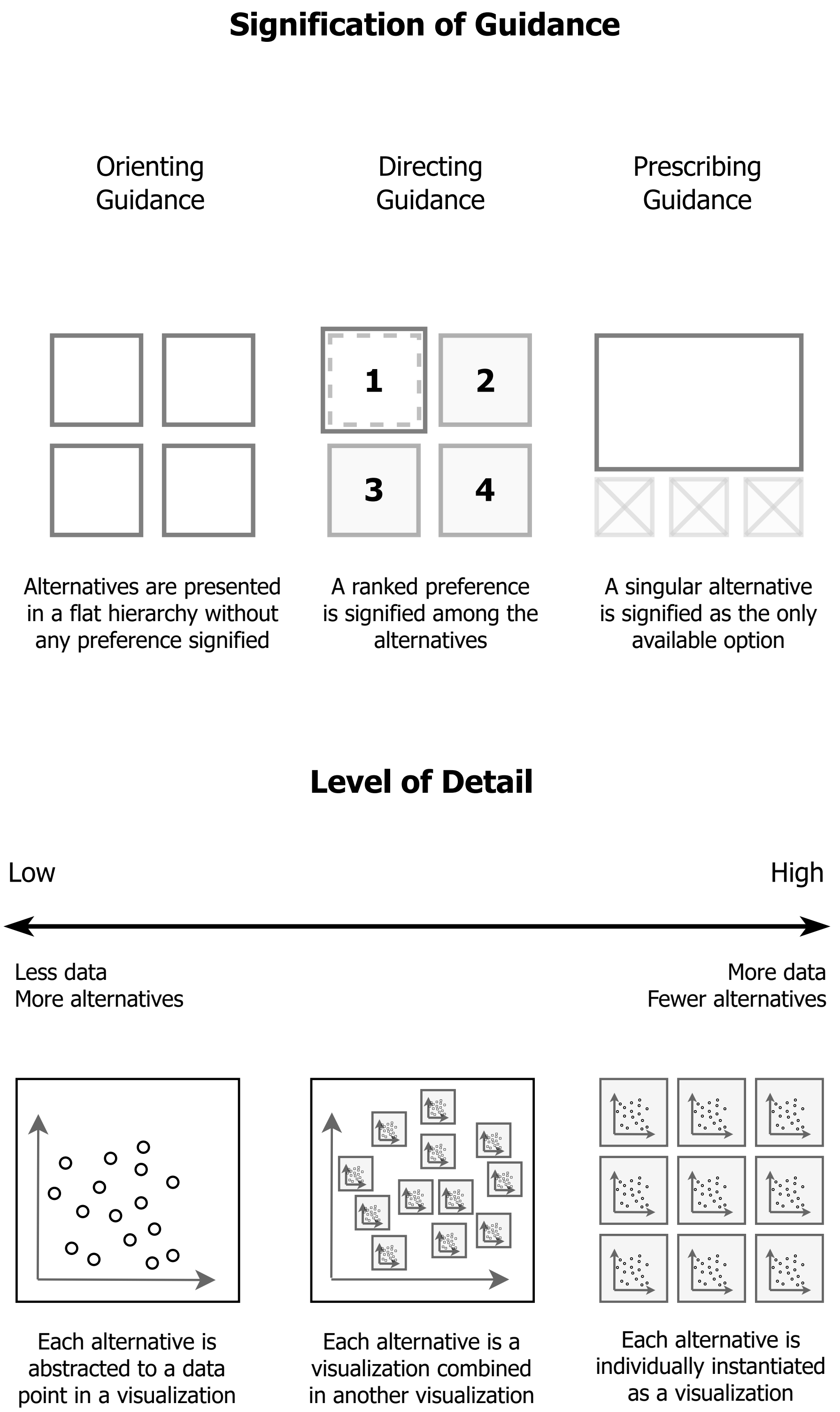


Figure 9.6: Abstracted illustrations of the guidance presentation according to varying significations of guidance and levels of detail.

The *level of detail* of each alternative can be influenced by many factors. For example, the available screen size and mental resource in context of use limits the level of details of presented alternatives. A higher number of alternatives may also limit how detailed each alternative can be shown. And the preferred alternative in directing guidance or the suggested alternative in prescribing guidance may be shown in greater detail. In the following, we discuss which options we have to accommodate different levels of detail.

- At the highest level of detail, the alternatives can be **individually instantiated** and then combined. This is often used to present a list or a grid of alternatives, similar to visualization spreadsheets. [46]
- When there are more alternatives, especially with some form of underlying relationships between them, their individual visualizations can be **combined into one visualization** – either a larger visualization, [167] or an ensemble visualization for showing general patterns and trends among the alternatives. [272]
- At the lowest level of detail, each alternative can be **abstracted to an individual data point**. The relevant metrics of each alternative can be abstracted in a single visualization or be directly encoded in the display of the alternatives using scented widgets. [277]

Additionally, these different levels of detail can also be combined to provide more adaptive and contextual guidance – the higher ranked or more important alternative(s) can be shown in greater detail with other alternatives abstracted into another visualization on the side, and more details about the abstracted alternatives can still be provided on demand when users hover over, zoom in, or select an alternative.

9.6.3 *Step 3.3 – Adapt to User Feedback*

9.6.3.1 *What*

In this final step, we consider how to adapt the provided guidance and its underlying MCDA model to user feedback to ensure the provided guidance is adaptive and controllable.

9.6.3.2 *Why*

The MCDA models in our guidance generation come with different means of criteria production and allow for different types of user input, enabling them to take user feedback into account. Using this possibility of adapting the guidance to the specific demands of data, task, and user fully enables the benefits of MCDA models.

9.6.3.3 *How*

As users interact with VA systems, many of their interactions can be recorded and analyzed as user feedback to adapt the provided guidance accordingly. To this end, the different components of guidance design must take into account the possibility of implicit (inferred) or explicit (input) user feedback. Making users aware of adaptations due to feedback is essential for them to make full use of the provided guidance – the occurrence and origin of an implicitly inferred adaptation should be communicated to users to avoid confusion, and the availability of direct control should likewise be signified.

Here we discuss three general perspectives on how guidance can be adapted to user feedback corresponding to the overall goals of the three stages of our method – considering whether or not guidance should be present as determined during the `intelligence` stage, the generation of guidance at the `design` stage, and the presentation of guidance at the `choice` stage.

- **Presence of guidance:** To ensure the provided guidance is non-disruptive, it should be possible for users to override the pre-determined need for guidance (Stage 1) by turning it off when not needed. Such mechanism can be directly accessed by the users through interactive elements on the user interface or prompted by user interactions.
- **Generation of guidance:** To provide adaptive and controllable guidance, the evaluation model pre-determined in Stage 2 can be interactively adapted. This can for example be achieved by including interaction metrics as criteria in the models themselves (e.g., the interaction history) [92] to automatically update the evaluation results based on user actions. Another option is for users to directly manipulate the models when they have the expertise to do so.
- To adapt the **presentation of guidance** to different contexts of use, the elements in the previous steps in this stage can also be made flexible – the guidance degree and detail level can be increased or decreased accordingly when users need more or less guidance. Such change can also be either made directly or inferred from the user interaction logs.

9.7 WORKSHEETS FOR APPLYING THE METHOD

Passing through the `intelligence`, `design`, and `choice` stages, we have provided a step-by-step method to design guidance for decision making in VA. In doing so, our method establishes a unified, reusable, and widely compatible guidance mechanism that generates transparent, adaptive, and consistent user guidance.

To provide concrete means that allow VA designers to put our method into practice, we developed a set of guidance design worksheets to accompany our method. We chose the format of worksheets, as they are accessible, flexible to be edited for different contexts, and well-suited for generating ideas and pen-and-paper prototypes especially in early design stages. Design worksheets are also a common in many aspects of visualization, such as visualization design, [174, 175] teaching and learning, [32] and creative ideation. [214]

To produce the worksheets, we went through a series of internal and external iterations to refine their design. In the beginning, we iterated on the worksheets three times internally. We started out by drafting the outline of each stage and step of our method on the worksheets. Then we went through a round of discussions that generated a list of improvements. After implementing them, we filled out the worksheets ourselves for a use case scenario of clustering analysis, and refined the worksheets based on this trial.

After the internal iterations, we went through a series of external iterations held in a workshop format. These workshops were conducted with three experts and lasted around 90 minutes each. Two of the experts were VA researchers with experience in designing VA systems, and the third was a researcher in Data Visualization and Human-Computer Interaction. The workshops were semi-structured and consisted of three parts: introduction, method walk-through, and follow-up questions. We started the workshops with a short introduction to the overall concept and examples of guidance in VA, the structure of our proposed method, and a basic example of MCDA. In the method walk-through, we first asked participants to identify the context of a VA system that they designed or used. Thereafter, they were asked to identify the decision points that were relevant to the users' workflows in the `intelligence` stage, develop a guidance generation mechanism for one important decision point in the `design` stage, and formulate the presentation of the alternatives in the `choice` stage. Finally, we ended the expert workshops with follow-up questions that we prepared and adapted with the observations from the method walk-through. In these questions, we focused on what could be improved in the worksheets and if the participants saw the benefits of using our method to formulate guidance.

Overall, the participants were all able to follow the steps in the worksheets, despite having some troubles articulating some of the key concepts. The two VA experts successfully identified a guidance problem in a VA system they previously designed and came up with

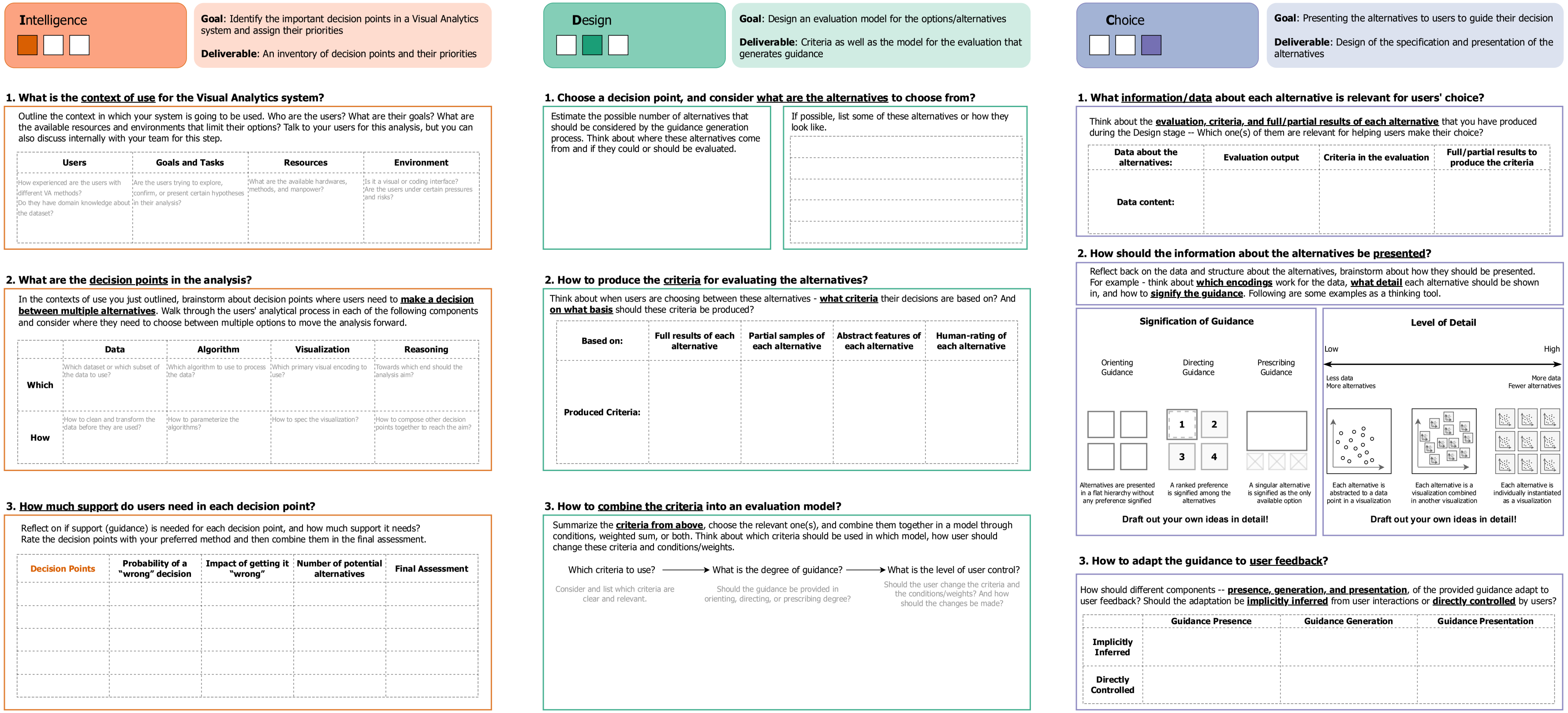


Figure 9.7: The worksheets accompanying our method correspond to its three stages, with three steps in each stage. Each worksheet consists of a title indicating the design progress, goal, and deliverable of the corresponding stage. For each step, the primary question to answer is presented with the keywords underlined. A prompt for each step outlines what the VA designers should achieve, and detailed instructions are provided to explain some of the key concepts.

a guidance solution with our worksheets, while the visualization expert was able to follow a set of pre-filled worksheets and provided some suggested improvements on the method and worksheets. In particular, they suggested to further clarify some of the key concepts in the method and emphasize the connections between the steps. Specifically, in the `intelligence` step, the meaning of the VA components – "data", "algorithm", "visualization", and "reasoning", as well as the types of knowledge gap – "which (scheme)" and "how (parameter)" can be challenging to delineate. We made revisions accordingly to produce our final worksheets. To clarify the key concepts, we added explanations to the worksheets. We emphasized the connections between the steps in the prompts to make the worksheets easier to follow. The final worksheets are shown in Figure 9.7.

9.8 Use case example and prototype

To provide a practical example, we present a use case based on our guidance method. To this end, we apply our worksheets on a scenario of cluster analysis and design an initial prototype through our method using existing visualizations and MCDA tools.

In the following, we describe how we apply each step in our guidance method in this use case, illustrate the developed prototype along the steps, and present the users' workflow after implementing the guidance. A set of filled worksheets was completed along the process and can be found at the end of this paper. For the prototype, we first drafted early iterations through pen-and-paper mockups and then implemented the prototype in Python. We utilized Bokeh for the visualizations, scikit-learn for the processing algorithms, and Scikit-Criteria for the MCDA methods.

9.8.1 *Stage 1 – Intelligence*

We start our guidance design through a context of use analysis. The *users* in our context of use are epidemiologists who have domain knowledge about the diseases under study and knowledge about different VA methods. Their *goal* is to explore and identify different types of patients for the same disease. To achieve such goal, the epidemiologists go through a series of *tasks* – they first clean the data and visualize them along some important features to first see if there are already some patterns in the patients. To further bring out these patterns, they then utilize dimension reduction and clustering algorithms to abstract the features and cluster the patients in different groups. During this process, they will need to experiment with different algorithms and parameters, then visually observe which set of results helps them to identify and determine the patient groups based on their expertise. As for *resources*, they have limited manpower and computing power. The *environment* for their analysis is not particularly time-sensitive and emergent. However, they tend to be less risk-taking, as the outcome of their analyses will inform future medical treatments and thus impact patients' lives.

With the context analyzed, we then go through the four components of the users' VA process in order to identify for which decision points guidance is needed. In the context of high-dimensional medical data, the epidemiologists would need to start with selecting the data dimensions to use and dealing with missing data that often appear in the medical context. With the data prepared, they then need to choose and parameterize the algorithms, such as dimension reduction and clustering, for grouping patients into different types. Finally, to uncover the patterns and insights from these computations, their results then need to be presented in visualizations with appropriate encodings and specifications to help epidemiologists visually observe these results and infuse their domain expertise. During these processes, the epidemiologists would also need to reason about how to combine the available data, algorithms, and visualizations together in order to uncover the insights they set off to seek. These decision points are summarized in Table 9.1.

Based on the context and the identified decision points, we then assess these decision points according to the three factors of guidance need on a scale of 0 to 3 – with "0" to signify the factor as "not important at all" and 3 to signify the factor as "extremely important". We multiply the factors and rank them from high to low to yield the final assessment (see Table 9.2). We chose to include 0 as the lowest point on the scale and to multiply the factors, as we consider if any factor in the assessment is 0 (not important at all), then there is no need for guidance at the corresponding decision point. In particular, different clustering algorithms and different numbers of clusters produce various stratifications of the patient cohort – each capturing a different property or insight into the disease. However, which clustering algorithms to explore and which sets of results are likely to contain reasonable stratifications are not known beforehand, and the epidemiologists would have to experiment with many of them to find the "needle in the haystack". This is where guidance becomes particularly helpful to point them in the directions of the most promising clustering results and achieve their goals of identifying patient types.

Table 9.1: A list of decision points we recognize in the use case of clustering analysis through the 4 components – data, algorithm, visualization, and reasoning, as well as 2 types – which (scheme) and how (parameter) that we discussed in Step 1.2.

Type	Data	Algorithm	Visualization	Reasoning
Which	Data dimensions to use	Dimension reduction and clustering algorithms to apply	Encoding of visualization to apply	Insights to uncover
How	Deal with missing data	Parameterize the algorithms	Spec the visualization	Combine other components

Table 9.2: Example inventory ranking the decision points's need for guidance for the use case of clustering analysis. We assessed the factors on a scale of 0 to 3 in their severity and multiplied the three factors to produce the final assessment.

Decision Points	Potential of Wrong Decision	Impact	Alternatives	Final Assessment
Clustering Algorithm (Which)	3	3	2	18
Data Dimension (Which)	2	3	3	18
Clustering Parameters (How)	3	3	2	18
Visualization Specifications (How)	2	2	3	12
Dimension Reduction Parameters (How)	3	2	2	12
Dimension Reduction Algorithms (Which)	3	2	1	6
Visualization Encoding (Which)	2	2	1	4
Missing Data (How)	3	1	1	3
Order of the Algorithms (How)	3	1	1	3
Insights to Discover (Which)	1	3	0	0

9.8.2 *Stage 2 – Design*

From the inventory of decision points, we have ranked the decision points and obtained the following three decision points with highest priority: clustering algorithms, parameters, and data dimensions. For the purpose of exemplifying our method, we focus on the two decision points of clustering algorithms and parameters, as these two decisions are interconnected and their guidance should be developed in conjunction.

The first step in this stage is to recognize the space of alternatives. The number of alternatives among the clustering algorithms is often not very large. For example, in the overview of clustering methods in the Machine Learning package for Python, scikit-learn, 11 algorithms are listed. [233] In our context of analyzing clusters of patients, the epidemiologists need to directly manipulate the number of clusters to explore the different resulting stratifications of the cohort of patients. Therefore, we decide to focus on the 5 algorithms with "number of clusters" as an available parameter – K-Means, Spectral Clustering, Ward Hierarchical Clustering, Agglomerative Clustering, and BIRCH. In our context, there are often not many different types of patients for the same disease, so the range for the number of clusters is also limited. Therefore, we set the range for the number of clusters from 2 to 10.

To produce the criteria, the evaluation of clustering algorithms and parameters are often based on similar metrics. In our case, the extrinsic measures based on a ground truth of actual class labels are not available. Therefore, intrinsic measures such as Silhouette Coefficient, Davies-Bouldin Index, and Calinski-Harabasz Index can be used to evaluate the separation and consistency of the clustering results. [141] These measures need to be calculated based on actual cluster results, but clustering the full dataset can take too much time due to the limited computing power. Therefore, we decide to produce these metrics with a small but representative sample of the full dataset. As we run these different algorithms on the sampled data, we can also measure their runtime.

To construct the MCDA model, we then consider the degree of guidance and level of control. As the users in our context have some knowledge regarding the algorithms and would like to explore alternative results, we decide to apply primarily directing guidance with additional orienting guidance to help. To provide directing guidance, the aforementioned criteria then need to be combined and some form of preference or recommendation should be derived from the evaluation. As our criteria are generated through computer-generated metrics, we first base it one preset criteria via functional approaches. The weights of each criterion were then set based on our experience with epidemiologists. Additionally, as the users in our context are able to understand the generated results and metrics, we then orient the users by presenting the clustering results and corresponding criteria from each algorithm and parameter. This helps to trace how the directing guidance was generated and orient users among the alternatives with detailed information. For allowing users to adapt the guidance to their specific needs, we present the underlying weights of each criterion and enable the users to manipulate the weights when they are not content with the provided guidance.

9.8.3 *Stage 3 – Choice*

To present the guidance, relevant data to be presented need to be recognized first. We present the produced results based on sampled data for each alternative, including the data points and their predicted class label. Furthermore, we present the meta-data including the evaluation criteria – specifically Silhouette Coefficient, Calinski-Harabasz Index, Davies-Bouldin Index, and runtime.

Next, we consider the degree of guidance and level of detail to produce the guidance presentation. For both of the decision points of clustering algorithms and parameters, we primarily apply directing guidance. Therefore, the preferences among the alternatives need to be signified accordingly. In the view of the produced results, we signify the ranking of the alternatives by highlighting the highest ranked one. Additionally, the evaluation table and figure of the algorithms and number of clusters also help to orient users among the alternatives by listing the alternatives and the evaluation results. For the level of detail, the number of alternatives for clustering algorithms is limited to the 5 algorithms that have "number of clusters" as a parameter. Therefore, they could all be presented in detail with the produced results and criteria. For the "number of clusters" parameter, we have 9 alternatives ranging from 2 to 10 clusters.We decide to visualize their evaluation results as a line chart on the side, where the users can hover over the data points to inspect the underlying criteria. The prototype with guidance is presented in Figure 9.8.

Finally, we adapt the provided guidance to user feedback. For the presence of guidance, we allow users to manually enable or disable the guidance elements (evaluation and ranking of alternatives), especially when users do not need to inspect or do not understand these detailed criteria. This also influences the presentation of guidance – with the evaluation table and figure that provide orienting guidance hidden, the system only provides directing guidance by indicating the highest ranked algorithm with its enlarged size.

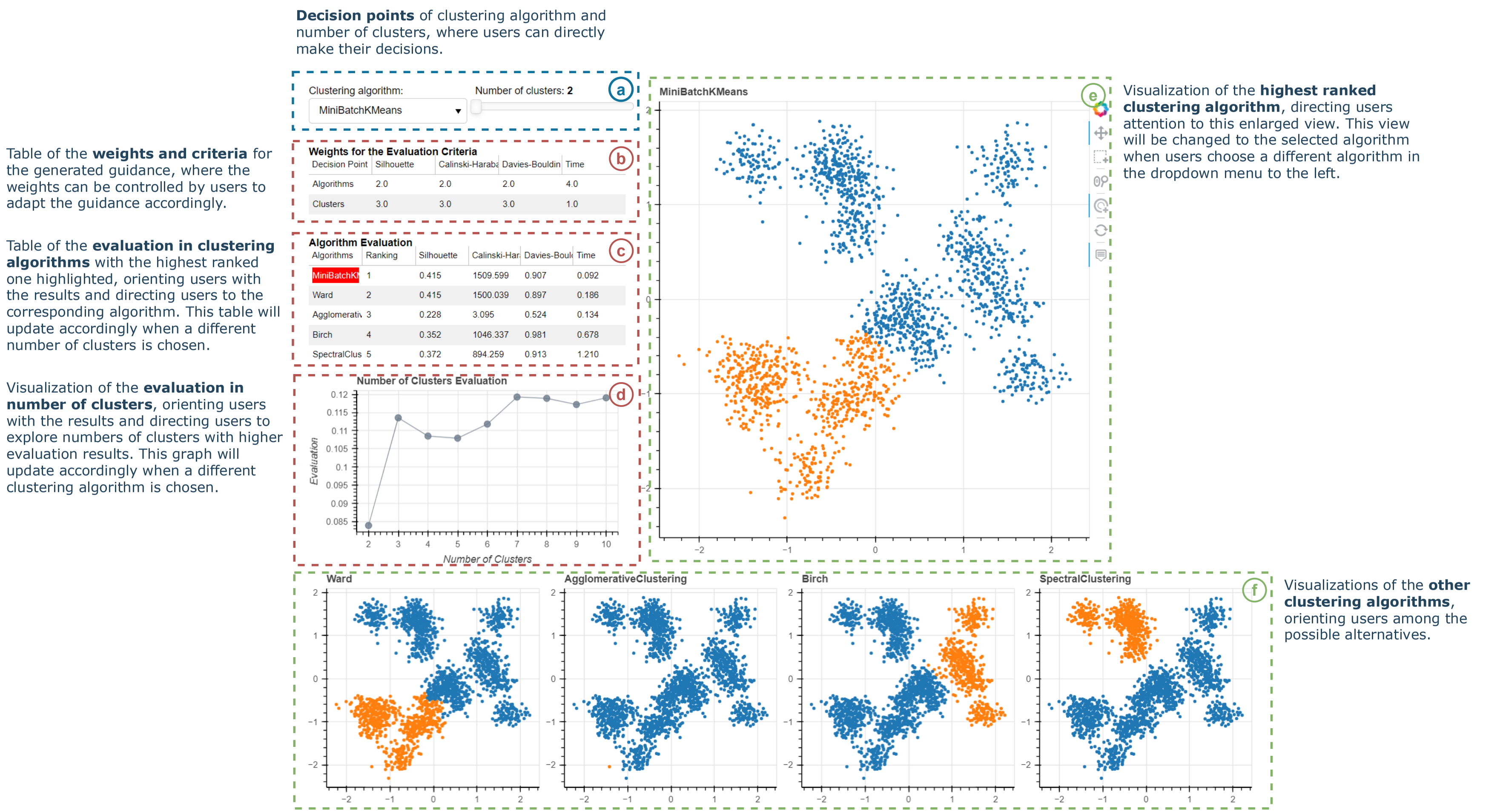


Figure 9.8: The annotated prototype for the use case with guidance produced based on our method, with the typical VA elements indicated in blue, guidance elements in red, and typical VA elements mixed with guidance in green.

9.8.4 *Use Case Example with Guidance*

After the guidance implementation, users enter their analysis with clear priorities and supporting information for the two decision points. For the algorithms, the enlarged view (Figure 9.8.e) and highlight color (Figure 9.8.c) guide users' attention to the highest ranked algorithm at the first glance. The detailed evaluation metrics (Figure 9.8.c) that generated the underlying guidance support the comparisons between the algorithms with important knowledge, while encouraging users to consider different alternatives. For the number of clusters, the evaluation chart (Figure 9.8.d) also indicates how this decision point might influence produced results with the selected algorithm, guiding users to compare and consider different numbers of clusters for their analysis. Moreover, as users change the number of clusters, the guidance is updated accordingly to encourage users to consider how different combinations of algorithms and numbers of clusters might impact their analysis. With the implemented guidance for the two decision points, we bring out several benefits to users that address important challenges they might encounter.

First, the direct presentation of alternatives (Figure 9.8.e/f) makes users aware of the decision points and their potential impacts on the results of the clustering analysis, encouraging users to explore the space of alternatives.

Second, the visual cues of color and size (Figure 9.8.c/e) indicate the highest ranked alternative, guiding users towards it. Meanwhile, how the underlying mechanism generated this ranking (Figure 9.8.b/c/d) is shown to support the trustworthiness of the provided guidance.

In addition, the presented visualizations (Figure 9.8.e/f) and evaluation criteria (Figure 9.8.b/c/d) allow the users to easily compare the alternatives, reason about them in a way that is grounded in the metrics, externalize their decision making process, and construct their evaluations, which in turn reduces their cognitive load. The possibility of manipulating the underlying criteria (Figure 9.8.b) enables users to adapt the guidance to their expertise and context of use, making the guidance design adaptive, flexible, and controllable.

9.9 DISCUSSION

To put our step-by-step design process in the context of existing research, this section discusses its commonalities and differences with four guidance frameworks that are closely related to ours:

- **Collins et al.'s 2018** paper collects a range of highly useful thoughts, arguments, models, and building blocks capturing guidance from its different perspectives of goals, requirements, roles, tasks, implementation, and evaluation. [51] The common theme underlying this paper is that of an "intelligent guide" or an "artificial intelligence-guided visualization".

- **Ceneda et al.'s 2020** paper introduces a framework for making design decisions on guidance functionality to be provided during visual-interactive data analysis. [37] It details aspects such as the requirements and goals of guidance, the knowledge gaps it addresses, the generation of guidance, and the users' feedback.
- **Pérez-Messina et al.'s 2022** paper proposes a typology of guidance tasks that connects the concept of guidance – its degrees (orienting, directing, prescribing) and the knowledge gap it addresses (target unknown, path unknown) – with user tasks (mainly search tasks). [203] This typology is not a design framework in itself, but it supports guidance design with a nuanced abstraction of guidance tasks.
- **Sperrle et al.'s 2022** paper describes a syntax to specify guidance functionality on the implementation level. [244] Its compact notation and low overhead allow for rapid prototyping of guidance, which makes it a good fit for iterative guidance design.

In the following, we will discuss these frameworks and our guidance design method with regard to common requirements and different contributions to guidance in VA.

9.9.1 *Common Requirements for Effective Guidance*

What makes for "good" guidance in Visual Analytics? Ceneda et al. state five requirements for "effective" guidance – *available, non-disruptive, adaptive, trustworthy*, and *controllable*. [37] Collins et al. includes similar requirements for "intelligent" guidance with different wording – *effective, adapted to the context (contextual), white-box,* and *right timing and mode* (see Figure 9.9). [51] Note that *effective* means in this case that the provided guidance should be "easily accessible" and "avoid distraction or obscuring the current visualization". This corresponds more closely to the requirements of *available* and *non-disruptive* from Ceneda et al., than to their overall goal of "effective" guidance.

A requirement that has received a lot of attention in guidance research lately is that of providing *adaptive* guidance, as it is called by Ceneda et al. This requirement appears under the name *contextual* guidance in Collins et al., emphasizing that guidance "should be *adapted* to the context of the user analysis process". Sperrle et al. cast

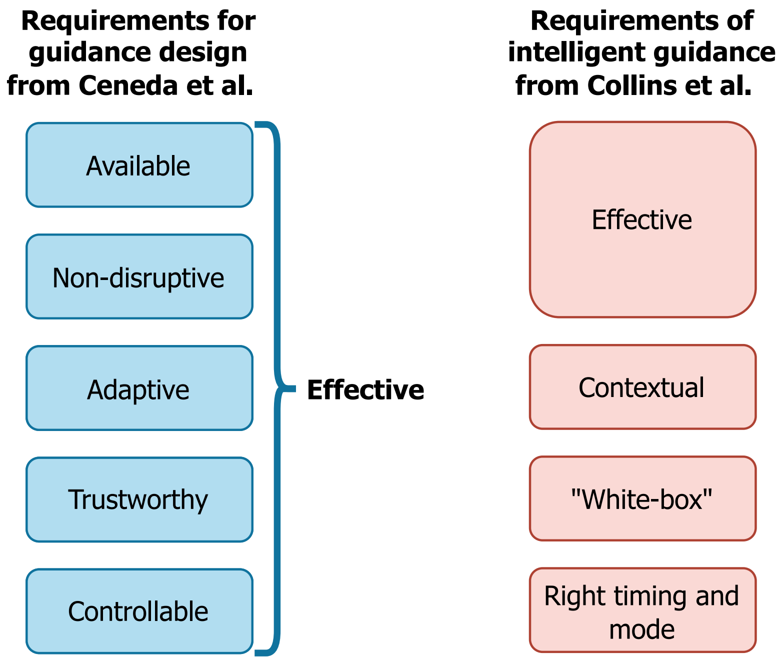


Figure 9.9: The requirements for guidance design from Ceneda et al. [37] (in blue) and Collins et al. [51] (in red).

this notion of adaptive guidance into a dedicated framework, [245] used it in complex guidance scenarios, [243] and built their recent guidance syntax around it. [244]

Furthermore, Ceneda et al. discuss the importance of *non-disruptive* guidance. This notion can also be found in Collins et al.'s requirement of providing guidance *at the right time and in the right mode*, which explicitly discerns between *synchronous guidance* that may disrupt and intervene with the analysis process, and *asynchronous guidance* that can be used or ignored as needed. The task typology by Pérez-Messina et al. makes prominent use of this distinction in their framework as well. [203] Interestingly, neither Collins et al. nor Pérez-Messina et al. connote *disrupting* or *synchronous* guidance necessarily as improper or inadvisable, but instead highlight the importance of making an informed and explicit choice about it.

9.9.1.1 *Our Method*

Many elements of the guidance design method presented here were purposefully included to meet these requirements. Through identifying and evaluating the decision points in a structured manner, we ensure the guidance is *available* when it is needed, yet *non-disruptive*. Furthermore, our underlying MCDA models can adapt to different guidance degrees and levels of user controls, making the produced guidance *adaptive*. As our method emphasizes the possibility of exposing the MCDA model to users, we also enable the produced guidance

Table 9.3: Comparisons of our work with selected papers on guidance design in VA. • indicates the corresponding work includes the element as a main focus, while ◦ indicates the work discusses the element without detailing specific processes or tools.

Paper	Contributions to VA Guidance		
	Conceptualizing	Designing	Implementing
Collins et al. (2018) [51]	•		◦
Pérez-Messina et al. (2022) [203]	•		
Ceneda et al. (2020) [37]		•	
Sperrle et al. (2022) [244]		•	•
our method	•	•	•

to communicate essential information about its *trustworthiness* and to act as a *white-box* through revealing the underlying mechanism. Finally, through adapting the guidance presence, generation, and presentation to user feedback, we further strengthen the ability of provided guidance to be easily *controllable*.

9.9.2 *Different Contributions to VA Guidance*

All mentioned frameworks address guidance from their own distinct perspectives and therein make different contributions to VA guidance that range from conceptual models to designing and ultimately implementing guidance. A summary of these contributions is given in Table 9.3.

9.9.2.1 *Conceptualizing Guidance*

From its inception, guidance in VA was likened to a car navigation system for visual-interactive analyses. [41] This metaphor is picked-up again by Collins et al., who point out the gap between this aspiration and the reality of current guidance functionality. [51] As the main issues causing this gap, they identify that existing guidance models – particularly Ceneda et al.'s original guidance characterization – are rather abstract and too far removed from practice to be useful. The paper then goes on to alleviate this issue by extending and detailing different aspects of guidance – e.g., task abstraction, user roles, implementation, and evaluation – to make the concept more actionable. In particular the aspect of task abstraction is then revisited in more detail by Pérez-Messina et al.'s typology. [203]

9.9.2.2 *Designing Guidance*

While the conceptual papers describe the tools for providing guidance (e.g., requirements, tasks, building blocks for implementation and evaluation), they do not detail how to use them to get from an identified knowledge gap to a suitable guidance solution. These more procedural concerns are addressed by Ceneda et al.'s 4-step framework for guidance designers in which they propose a sensible order in which to make the different necessary considerations for arriving at guidance solutions. [37] As this 4-step framework aims to anticipate user problems at design time, it has only limited possibilities to adapt to changing user needs as they emerge at runtime during exploratory, open-ended analysis sessions. Consequently, Sperrle et al. call this a *theoretical design process* and instead propose a *strategy-centered guidance design* in which various different guidance strategies are developed at design time and then chosen dynamically as needed at runtime. [244]

9.9.2.3 *Implementing Guidance*

Having a guidance design – i.e., all questions regarding the why, when, what, and how of guidance are answered – still does not realize the thus specified guidance in code. To that end, Collins et al. provide a high-level discussion on guidance implementation (e.g., sources of information, computational processes, interaction modalities) using an input-compute-output structure. [51] These can help to further detail the design considerations into implementation considerations, but it still does not lead to actual code. The framework from Sperrle et al. is the first practical contribution to tackle this last mile bridging specification to implementation of VA guidance. [244] By expressing their guidance strategies through a declarative grammar, they are able to automatically generate guidance functionality from the specification.

9.9.2.4 *Our method*

As can be seen, all existing frameworks have a particular focus in their contributions to VA guidance, whereas our method contributes to all three aspects:

- **Conceptualizing Guidance:** Our method reframes and extends the common notion of "guidance for knowledge gaps" to "support for decision points".
- **Designing Guidance:** This new perspective on guidance allows us to re-structure the guidance design process along the decision making process.
- **Implementing Guidance:** It also allows us to use existing decision support systems – in our case MCDA models – as generic way to generate guidance by evaluating alternatives.

In this way, our method provides an end-to-end solution from a way of thinking about guidance to a matching way of designing guidance all the way to its implementation. To bridge the gap between theory and practice, we provide the worksheets as practical tools to be used for guidance design and we root our method in the MCDA approach for which a wide range of libraries and packages in different programming languages are available to jumpstart guidance generation without much additional overhead.

That being said, our method is not necessarily ideal for designing guidance in all cases. In scenarios where the evaluation criteria for an MCDA model are hard to explicitly produce, the framework from Ceneda et al. might work better as a first design iteration. [37] Meanwhile, the compact format of the guidance strategy template from Sperrle et al. might be more suitable in fast-paced iterations of guidance design. [244]

9.10 CONCLUSIONS

In this paper, we introduced a generic guidance design method through decision support. To this end, we re-framed guidance as supporting decision points, presented and detailed our generic guidance method through three steps for each of the three stages in the decision making process, and produced a set of flexible worksheets through internal iterations and expert workshops based on our method. Additionally, we developed an initial prototype with the produced worksheets to exemplify the usefulness and applicability of the method and discussed our method in the context of existing general research on VA guidance. With both theoretical and empirical insights, this work expands the concept of guidance in VA with decision making research and provides a practical guidance method covering the end-to-end process of producing guidance in VA.

As for future work, in particular the discussions from the previous section make it obvious that a "nested model of guidance design" akin to Munzner's nested model for visualization design is almost within reach. [187] By applying the different frameworks at the different levels of such a nested model – e.g., Pérez-Messina et al.'s task typology for the task abstraction and Sperrle et al.'s Lotse guidance library or our MCDA models at the algorithm level – it should be possible to realize a nested design process specifically for VA guidance. Likewise, different aspects of Collins et al.'s considerations for evaluating guidance could then be related as validation measures to the different levels. For example, their suggestion to do user studies with "complex information seeking tasks" could be used to validate the guidance design, while the "metrics for automated monitoring" could be useful to benchmark the algorithmic implementation of the guidance generation. Filling in the still missing pieces to complete this picture – e.g., a data abstraction for guidance in VA to complement the existing task typology – pose formidable research challenges in this direction.

Intelligence

Goal: Identify the important decision points in a Visual Analytics system and assign their priorities

Deliverable: An inventory of decision points and their priorities

1. What is the <u>context of use</u> for the Visual Analytics system?

Outline the context in which your system is going to be used. Who are the users? What are their goals? What are the available resources and environments that limit their options? Talk to your users for this analysis, but you can also discuss internally with your team for this step.

Users	Goals and Tasks	Resources	Environment
Epidemiologists who have domain knowledge about the diseases under study and knowledge about different VA methods	Explore and identify different types of patients for the same disease. Data cleaning, processing, and finding patterns.	Limited manpower and computing power.	Visual interface More possible risks, as medical contexts could influence patients' lives

2. What are the <u>decision points</u> in the analysis?

In the contexts of use you just outlined, brainstorm about decision points where users need to **<u>make a decision between multiple alternatives</u>**. Walk through the users' analytical process in each of the following components and consider where they need to choose between multiple options to move the analysis forward.

	Data	Algorithm	Visualization	Reasoning
Which	Which dimensions of the data to include?	Which dimension reduction and clustering algorithm to use for processing the data?	Which primary visual encoding to use to find the clusters?	Which type of clusters/patterns to discover -- e.g., to find different mechanisms that cause the same disease, or
How	How to clean up messy entries and make placeholders for the missing data?	How to parameterize the algorithms? e.g. number of clusters	How to specify the visualization to bring out the patterns more clearly?	How to combine different data dimensions and algorithms according to the reasoning -- e.g. how to combine dimension

3. <u>How much support</u> do users need in each decision point?

Reflect on if support (guidance) is needed for each decision point, and how much support it needs?
Rate the decision points with your preferred method and then combine them in the final assessment.

Decision Points	Probability of a "wrong" decision	Impact of getting it "wrong"	Number of potential alternatives	Final Assessment
which data features/dimensions	medium	high	medium	2 x 3 x 2 = 12
which clustering algorithm	**high**	**high**	**medium**	**3 x 3 x 2 = 18**
order of the algorithms	high	medium	low	3 x 2 x 1 = 6
which primary encoding	medium	medium	low	2 x 2 x 1 = 4
...	...	...	...	...

Design

Goal: Design an evaluation model for the options/alternatives

Deliverable: Criteria as well as the model for the evaluation that generates guidance

1. Choose a decision point, and consider what are the alternatives to choose from?

Estimate the possible number of alternatives that should be considered by the guidance generation process. Think about where these alternatives come from and if they could or should be evaluated.

Scikit-learn has a handful of clustering algorithms -- around 10 were listed in their documentation for comparisons

If possible, list some of these alternatives or how they look like.

- k-means
- DBSCAN
- hierarchical clustering
- Spectural clustering
- birch

2. How to produce the criteria for evaluating the alternatives?

Think about when users are choosing between these alternatives - **what criteria** their decisions are based on? And **on what basis** should these criteria be produced?

Based on:	**Full results of each alternative**	**Partial samples of each alternative**	**Abstract features of each alternative**	**Human-rating of each alternative**
Produced Criteria:	Full results might be too computationally expensive	Silhouette Coefficient Davies-Bouldin Index Calinski-Harabasz Index Rand index and mutual information based score can be used when the ground truth is available	Computation time	Subjective quality

3. How to combine the criteria into an evaluation model?

Summarize the **criteria from above**, choose the relevant one(s), and combine them together in a model through conditions, weighted sum, or both. Think about which criteria should be used in which model, how user should change these criteria and conditions/weights.

Which criteria to use? → What is the degree of guidance? → What is the level of user control?

Which criteria to use?	What is the degree of guidance?	What is the level of user control?
Consider and list which criteria are clear and relevant.	Should the guidance be provided in orienting, directing, or prescribing degree?	Should the user change the criteria and the conditions/weights? And how should the changes be made?
Silhouette Coefficient Davies-Bouldin Index Calinski-Harabasz Index Computation time	Mostly directing Possibly some orienting	Users are experts, so they might want to change the criteria weights to their liking

Choice

Goal: Presenting the alternatives to users to guide their decision

Deliverable: Design of the specification and presentation of the alternatives

1. What <u>information/data</u> about each alternative is relevant for users' choice?

Think about the **<u>evaluation, criteria, and full/partial results of each alternative</u>** that you have produced during the Design stage -- Which one(s) of them are relevant for helping users make their choice?

Data about the alternatives:	**Evaluation output**	**Criteria in the evaluation**	**Full/partial results to produce the criteria**
Data content:	Ranking of the alterntaives	Scores in the criteria: Silhouette Coefficient Davies-Bouldin Index Calinski-Harabasz Index Computation time	Partial results of each alternative are also available as labels for each patient

2. How should the information about the alternatives be <u>presented</u>?

Reflect back on the data and structure about the alternatives, brainstorm about how they should be presented. For example - think about **<u>which encodings</u>** work for the data, **<u>what detail</u>** each alternative should be shown in, and how to **<u>signify the guidance</u>**. Following are some examples as a thinking tool.

Signification of Guidance

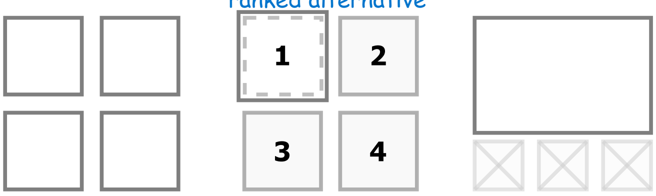
Orienting Guidance

For presenting the criteria information

Directing Guidance

Using color and size to indicate the highest ranked alternative

Prescribing Guidance



Alternatives are presented in a flat hierarchy without any preference signified

A ranked preference is signified among the alternatives

A singular alternative is signified as the only available option

Draft out your own ideas in detail!

Level of Detail

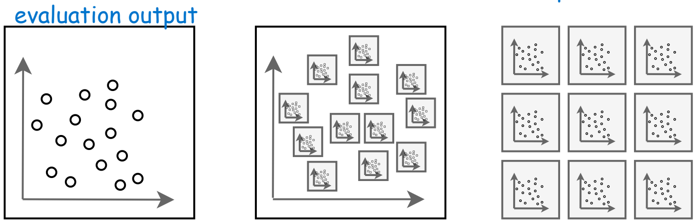
Low — High

Less data
More alternatives

For the criteria and evaluation output

More data
Fewer alternatives

For the partial results



Each alternative is abstracted to a data point in a visualization

Each alternative is a visualization combined in another visualization

Each alternative is individually instantiated in a visualization

Draft out your own ideas in detail!

3. How to adapt the guidance to <u>user feedback</u>?

How should different components -- **<u>presence, generation, and presentation</u>**, of the provided guidance adapt to user feedback? Should the adaptation be **<u>implicitly inferred</u>** from user interactions or **<u>directly controlled</u>** by users?

	Guidance Presence	**Guidance Generation**	**Guidance Presentation**
Implicitly Inferred	Possibly turn off the guidance when users do not follow the suggestions	Users' preferences among the alternatives can be inferred	Similar to presence -- the details can be cut down when users do not interact with them
Directly Controlled	User should be able to turn off some of the guidance	The criteria and weights can be changed	Details of the evaluation criteria can be hidden when not needed

10

PAPER D: VIBROTACTILE GUIDANCE

Exploring Vibrotactile Cues for Interactive Guidance in Data Visualization

Wenkai Han and Hans-Jörg Schulz

Department of Computer Science, Aarhus University, Denmark

ABSTRACT

In visualization, user guidance has become an essential concept to aid users in making informed decisions ranging from what subsets to focus on in the data space to which regions to explore in the view space. To guide users, predominantly visual cues like colors or arrows are used to indicate particular targets or directions. In this paper, we explore the possibility of another sensory channel for guidance cues: vibrotactile feedback. To that end, we explore different properties of the vibrotactile channel (e.g., amplitude and duration) and discuss their potential use as guidance cues. We then report on an experiment (N=14) in which we investigate possible vibrotactile cues in comparison to visual cues and to a combination of visual and vibrotactile cues for a guided selection scenario and a guided navigation scenario. Although none of the vibrotactile cues significantly outperformed the visual cues, our study results shed light on a number of practical issues when using vibration for user guidance – including differences between various types of vibrotactile feedback, as well as diverging performance for different guidance scenarios.

10.1 INTRODUCTION

Over the last years, guidance has emerged as an important tool to facilitate user interaction and decision making in visual data analysis. Ceneda et al. define guidance in the context of visual analytics as *"a computer-assisted process that aims to actively resolve a knowledge gap encountered by users during an interactive visual analytics session"* [41]. Guidance can be provided in a variety of ways, depending on the type of the knowledge gap between the users and their goal and the desired level of computer assistance [38, 40, 51]. While the variety of guidance explicitly includes non-visual means of guidance, subsequent research has so far exclusively focused on visual guidance cues and left out any other type of cues (e.g. haptic and sonic ones) [39, 41].

This stands in contrast to a growing interest in using non-visual forms to represent data: data physicalization approaches explore haptics to make data tangible [124], data sonification uses sound to make data audible [133], and data olfactation applies scents to make data perceivable through smell [202].

These non-visual forms of data presentation not only solve problems in traditional visual presentation but also provide additional benefits to the existing visualization systems. In visual analytics tasks where users are already visually overloaded, alternative forms of data presentation can help to solve visual conflicts and ease the visual load. Additionally, using non-visual forms in data visualization can also aid visually impaired users and provide an immersive experience for others [63, 101].

Among non-visual presentation forms, vibrotactile approaches have proven as only second to visual ones with respect to sensory bandwidth [56, 140, 143]. In the light of input devices with vibrotactile feedback now being readily and commercially available (e.g., gaming mice, Microsoft Surface Dial), we explore the use of vibrotactile means for guidance in data visualization. Among the different options like amplitude or frequency of a vibration, we identify possible vibrotactile cues that seem to be good candidates for guiding users in interactive visualizations. We put these cues to a test in a user experiment that compares performance and experience between visual cues, vibrotactile cues, and a combination of both using an off-the-shelf vibrating mouse.

As a result of these experiments, our work not only proposes and validates a new vibrotactile channel for guidance in data visualization, but also opens up a new design space of different vibrotactile patterns for interaction design in data visualization. The analysis of the gathered user performance and experience data from different vibrotactile patterns provides further insights in how such guidance should be designed.

In the following, we present the related work on vibrotactile data presentation. Thereafter, we present the setup of our experiment on two data visualization scenarios and report the results of the experiment. From our findings, we present open research questions for vibrotactile guidance in visual analytics.

10.2 VIBROTACTILE DATA PRESENTATION

Although nearly untouched in the area of guidance, vibrotactile cues have been well studied for presenting data in the existing literature. In general, vibrotactile data visualization has been proven useful in presenting ordinal [256] and categorical data such as discrete directions and locations [127, 144]. But, as detailed in the following, different vibration parameters lend themselves to the presentation of data to different degrees. For example, frequency and amplitude have been shown to represent relative values better than absolute ones [90, 256].

10.2.1 *Amplitude*

Amplitude refers to the intensity of vibration. It can be described in G as acceleration or dB as acceleration level. The recommended range of amplitude has been discussed with regard to the acceleration level. Craig and Sherrick [57] found that 28 dB would be a practical maximum of vibration, as human perception deteriorates above this threshold. Gunther [99] reported that vibration of more than 55 dB might invoke pain and should be avoided.

Even though a vibration's amplitude is continuous, it is not well suited for representing continuous data. This is because it can be rather hard for human beings to differentiate between different intensities of vibrations. Gill [90] states that no more than four different intensities should be used.

Early studies [127–129] on using vibrotactile and auditory cues to present bivariate and trivariate maps of categorical data suggested that the vibrotactile cues with different amplitudes can increase the recall rate of the information, while the completion time is similar to visual cues. However, they also pointed out that the vibrotactile cues might interfere with other forms of cues for presenting information.

10.2.2 *Frequency*

The frequency of a vibration refers to the number of times that the vibration unit goes back and forth between the amplitude per unit of time. The range of frequencies that human beings are most sensitive to is between 20 Hz and 1000 Hz, with the optimal point lying at around 250 Hz [99].

The change of frequency can usually be used to encode numerical changes similar to amplitude. However, a lower frequency of vibration would provide a more loose and rough feeling to the users, while a higher frequency would invoke a tighter and finer sensation.

Like amplitude, the number of different frequencies that can be differentiated is limited. No more than 9 different levels of frequency should be used, and the difference between the levels should be no less than 20% [90]. For representing ordinal information haptically through different frequencies, a study with a force-feedback rotary device found that the response time for using vibrotactile cues is significantly better than for positional cues, but the accuracy can be 10% to 20% lower [256]. However, the frequency of vibration might interfere with other parameters like amplitude. Enriquez and MacLean [163] suggest that the frequency range of 5 Hz to 20 Hz would reduce this interference and increase the expressive capability of the vibrations.

Different rates of the change in amplitude and frequency can also carry more intricate meanings. Participants in a study by Gunther [98] describe an abrupt change in these parameters as similar to a tap against the skin, while a gradual change feels like something rising up or out of the skin. In addition, the profiles of these changes (e.g. linear, Gaussian, or polynomial profiles) can have an influence on user performance and experience. A study on haptic feedback found that linear patterns might lead to lower performance and are less preferred by users than non-linear ones [14].

10.2.3 *Waveform*

Sine, square, triangle and sawtooth waveforms are the most commonly used waveforms of vibrations. Although people are relatively good at discerning between two different waveforms [98], studies on discriminating between more than three waveforms have been lacking. Rovan and Hayward [219] explain that the sine waveform is commonly used as it provides a sense of smoothness, while other waveforms like square and sawtooth would be rougher.

Different from amplitude and frequency, the change of waveform can be used to imply a change between ordinal or categorical data values. As the different waveforms have different degrees of smoothness, it would be possible to use it for representing discrete data. However, we should also be aware that human beings' ability to differentiate between more than three waveforms has not been proven. Therefore, caution is needed when using waveforms to encode data values.

10.2.4 *Duration*

Different duration of vibrotactile cues can encode information by changing the length of vibration. Stimuli less than 100 ms are usually perceived as a tap [98], while longer stimuli combined with different patterns can deliver a wide range of physical perceptions.

The change of duration and intervals of vibrations can be used to combine individual vibrations into sequential patterns. These patterns can be used to encode data similar to Morse code, but with the potential of adding in other properties. For example, a sequence of a shorter interval with strong vibrations can suggest a more intense feeling, and vice versa. Compared to encoding such information only with the intensity, we argue that the metaphor for a value or the severity of a situation might be more easily understood by users through such sequences.

10.2.5 *Pattern*

The pattern of a vibration refers to the change in amplitude, frequency, waveform, and duration. These patterns can encode information using metaphorical vibrations (the single tap, multiple knocks, a faint buzz, etc.) or not.

Tactons, a design structure for vibrotactile patterns, in particular, are used in various studies to represent different information [25, 28, 29, 208]. Tactons are constructed with different vibrotactile patterns to non-visually communicate complex concepts by compounding different parameters such as frequency, amplitude and duration [25]. For example, a gradual increase in amplitude can represent the process of user actions, and a click vibration at the end can indicate the action has been successfully executed. Studies on different tactons all suggested that vibrotactile cues have a rather good identification rate [5, 28, 29].

10.3 A USER STUDY ON VIBRATION AS A GUIDANCE CUE

In the real world, vibration is frequently used to guide humans. Examples range from laser pointers that signal through vibration that the presentation time is almost up, to rumble strips being used as road markings to draw drivers' attention to potential dangers. Hence, it is only reasonable to adapt this tried and true idea for visual exploration, so as to see if it works for this domain as well. This makes particular sense, as vibrotactile feedback has been made available in several commercial input devices. In Table 10.1, we present how different commercially available devices support the different customized parameters for vibrotactile cues. Note that in this table, the degree of support is defined by their application programming interfaces (APIs) as of October 2020. To adapt vibration for visual exploration, we went through internal discussions and pilot testing to build a software prototype and to plan a user experiment to observe vibration-guided visual data exploration.

Table 10.1: A List of Widely Available Devices with Input Capability and Vibrotactile Output

Parameter & Device	**Amplitude**	**Frequency**	**Waveform**	**Duration**	**Pattern**	**Source**
SteelSeries RIVAL mouse	5 variations	No	No	Yes	Limited Selection	[250]
Microsoft Surface Dial	Yes	No	No	Yes	Limited Selection	[180]
Apple Magic Trackpad 2	Yes	Yes	No	Yes	Yes	[120, 121]
MacBook Pro Trackpad (2016 or later)	Yes	Yes	No	Yes	Yes	[120, 121]
iPhone (iOS 13 or later)	Yes	Yes	No	Yes	Yes	[120, 121]
Apple Watch	Yes	Yes	No	Yes	Yes	[120, 121]
Mi Band 4/5	3 variations	No	No	Yes	Limited Selection	[260]

10.3.1 *Participants*

14 participants (6 female and 8 male) were invited to a one-hour experiment session. Participants were students and researchers across different faculties from Aarhus University. The age of the participants ranged from 22 to 40. Among them, two were left-handed but used their right hands for computer mice in daily life. Two of the male participants were red-green colorblind. Two female participants had myopia, but their eyesight was corrected by their glasses. None of the participants had prior experience with haptic feedback in data visualization or was familiar with our prototype.

10.3.2 *Apparatus*

The experiment was conducted on a 14-inch laptop (HP Elitebook 840 G5) with Intel Core i7-8550U at 1920 pixels $\times$ 1080 pixels resolution. The mouse was a SteelSeries RIVAL 710 with embedded vibrotactile feedback. Participants interacted with the prototype through the mouse with their right hand. A Logitech G240 cloth mouse pad was used to provide a consistent surface for the mouse movement and vibrotactile feedback.

The choice for a vibrating computer mouse was made based on several reasons. First, as most current interactions with data visualization are still performed through computer mice, using a commercially available mouse makes our design more accessible for different users without the need to learn how to interact with an unusual device. Second, as the mouse has embedded vibrotactile feedback, users do not need to change between two different devices during the experiment (e.g., a mouse and a vibrotactile device). Finally, an off-the-shelf computer mouse makes it simpler to reproduce and extend our work.

The tactile motor in the mouse was an ELV1030A Linear Resonant Actuator (LRA) from AAC Technologies. The maximum vibration acceleration level was at 1.7G, and the resonant frequency was set to 205 Hz. The waveforms we used were predefined in the SDK. We chose the "ti predefined buzz" group of vibrotactile cues, as they had more constant waveforms and more varieties of amplitudes. The selection of vibrotactile cues can be found on the SDK's webpage [250].

10.3.3 *Tasks and Visual Cues*

There are two types of guidance in visual analytics based on the types of the knowledge gap [39–41]. For situations in which the *target is unknown*, a user knows how to reach a certain target, but does not know what that target is. Whereas for situations in which the *path is*

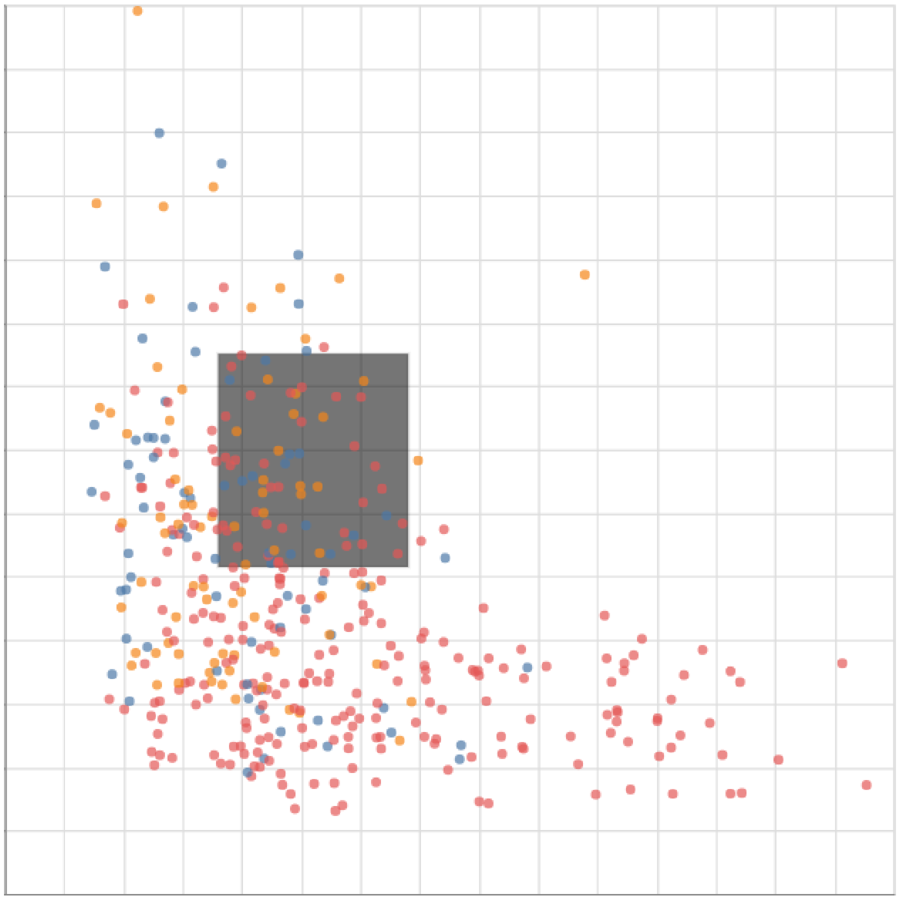


Figure 10.1: Example of the Prototype in the Selection Task

unknown, a user knows the target, but does not know how to reach it. We aim to cover both of the two types in our study by matching them with two of the most prominent visualization tasks: selection and navigation.

10.3.3.1 *Selection Task*

Selecting data items in a plot is one of the most common actions in visual data exploration [278], with the rectangular brush being probably the most popular interactive tool to do so. Yet, how many items should one select? This question stands at the heart of visual analytics approaches, such as the one presented by Angelini et al. [2], where too large a selection renders the resulting computation intractable due to combinatorial explosion, while too small a selection renders the result of that computation statistically insignificant. And even if one knew the optimal number to select, how could one judge if one was close to that number – in particular in the presence of overplotting and visual clutter? This scenario presents a case of an *unknown target*, where users are unaware of the specific target, but guidance can direct users. Following this idea, we built a prototype where the desired number

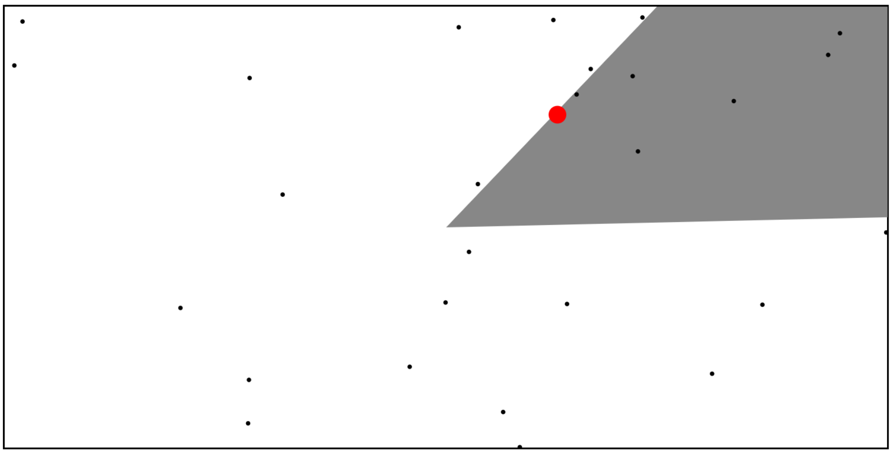


Figure 10.2: Example of the Prototype in the Navigation Task

of selected items would be hinted at by the background color of the brush (see Figure 10.1). When the number of selected items is closer to the desired number, the intensity of the background color would increase, and vice versa.

10.3.3.2 *Navigation Task*

The most universal form of navigation is panning. Whether one peers over a map or scrolls through a document, panning is embedded in many routine tasks. Hence, we use panning for testing navigational guidance cues that are inspired by the look-ahead radar view [259]. In the look-ahead radar view, an arc will appear when users are panning a graph visualization in the direction in which off-screen items of interest lie. Similar to the design in the selection task, we are also using the color opacity of the arc to indicate the speed with which the current panning movement is closing in on the target (see Figure 10.2). The highest speed is achieved by moving directly towards the target, lower speeds result from moving in its general direction, and "negative speed" occurs when moving away from it. With increasing speed, the opacity of the background color also increases, and vice versa. The navigation task corresponds to the *path unknown* type of guidance where users know the target they are searching for – in our case a red dot – but as the target is out of sight, the users are unaware of the path to that red dot.

10.3.3.3 *Measures against Learning Effect*

To reduce the influence of potential learning effects, in both of the tasks, the visualizations and the type of cue were randomly generated for each trial. The data in the selection task were based on the cars dataset by J. C. Schlimmer obtained from the UCI Machine Learning Repository [70]. We randomly scaled each data point from 0.5 to 1.5

times of their original values for each trial. As such, the visualization will be different for each trial. This will prevent participants from adopting the same strategies for selecting data points. For the navigation task, the data points were randomly generated using a normal distribution function. μ was set 10 times of the viewport size, and σ was set to 150. The parameters were chosen to minimize the possibility that users might pan through empty areas.

Additionally, for the selection task, we selected three different selection targets (15, 60, or 240 points) to avoid participants learning a fixed strategy and to cover different scenarios. Similarly, we also selected three distances of the target point from the starting point (2400, 4000, or 6000 pixels) for the navigation task.

10.3.4 *Vibrotactile Cues*

As covered in Section 10.2, several parameters can be changed for vibrations: amplitude, frequency, waveforms, etc. Among them, amplitude is the most commonly used one in previous research due to its rather high expressiveness [57, 90, 99]. Furthermore, as we change the opacity of color for our visual cue, a parameter of vibration that is similar to the color opacity would make the different cues more comparable in our study. Thus, the change of amplitude, which is usually characterized as the intensity of vibration, was chosen as the main focus.

Through the SteelSeries GameSense SDK of the used mouse, we could set 5 different amplitudes for each vibration pattern. As suggested in Section 10.2, a non-linear profile usually provides better user performance and experience, and we therefore fitted the 5 different amplitudes to a polynomial curve in our prototype. Figure 10.3 shows an example of such curve for the increasing vibrotactile cues in the selection task with the goal of 240 targets based on a quadratic formula, where the number of selected points x and vibration amplitude y have a relationship of $x = 2.4 \times (y \div 10)^2$ before the highest point, then it has a reversed pattern after the highest point.

During the development and pilot testing of the tasks, we also realized that even with a non-linear curve for the increment of vibration amplitude, it was hard to tell which vibration amplitude is the highest. Therefore, we also tested out a decreasing pattern with the same curve as the increasing one. A threshold pattern, where the vibration will only be triggered when hitting the right target/path, was added, as it provides a discrete and thus more accurate cue compared to continuous ones.

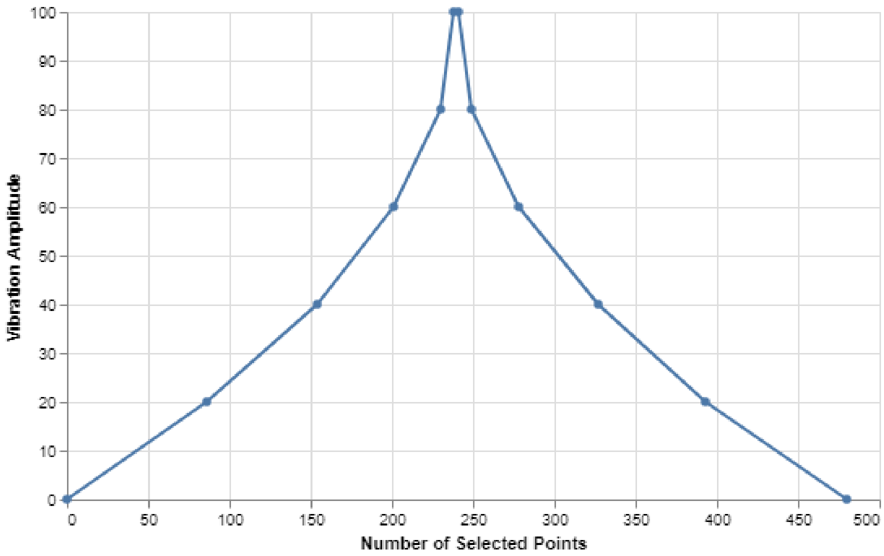


Figure 10.3: The Polynomial Curve for Increasing Vibrotactile Cues in Selection Task with the Goal of 240 Targets

10.3.5 *Type of Cues for Guidance*

Combining the color cues and vibrotactile cues, seven types of cues were tested for guidance in our experiment. These were color, increasing vibration, decreasing vibration, threshold vibration and combinations of color with the three vibrations respectively.

10.3.5.1 *Color (C)*

For color cues, we used a gray (#808080) visual cue with different opacity to guide a user's action. The opacity was generated using a polynomial curve. We used opacity instead of hue for the visual cue, as we found in the pilot study that it is more comparable to the amplitude parameter we use for the vibrotactile feedback. Participants in the pilot study stated that both opacity and amplitude denote the intensity of the corresponding feedback.

In the selection task, the color of the brushed background indicates how close the number of brushed points is to the target number. The closer the number of brushed points is to the target, the darker (higher opacity) the color becomes. In the navigation task, an arc following the panning direction indicates how fast the cursor is moving towards the target. The faster the cursor is moving towards the target, the darker its color.

While there are plenty of other visual cues one could use, every added cue significantly increases the number of possible combinations to test during our user study. As the focus of our study is on different vibrotactile cues, we only chose one visual cue based on the feedback we got in the pilot study.

10.3.5.2 *Increasing Vibration (IV)*

For increasing vibration, a polynomial pattern was applied for the amplitude of vibration. In the selection task, the increasing vibration indicates how close the number of brushed points is to the target number. The closer the number of brushed points is to the target, the higher the amplitude. In the navigation task, the increasing vibration indicates how fast the cursor is moving towards the target. The faster the cursor is moving towards the target, the higher the amplitude.

10.3.5.3 *Decreasing Vibration (DV)*

For decreasing vibration, a reversed polynomial pattern was applied for the amplitude of vibration. In the selection task, the decreasing vibration indicates how close the number of brushed points is to the target number. The closer the number of brushed points is to the target, the lower the amplitude. In the navigation task, the decreasing vibration indicates how fast the cursor is moving towards the target. The faster the cursor is moving towards the target, the lower the amplitude.

10.3.5.4 *Threshold Vibration (TV)*

The third vibrotactile cue uses a threshold for triggering the vibration.

In the selection task, the highest amplitude of vibrotactile feedback is provided only when participants select the target amount of points with an error margin of one. In the navigation task, the highest amplitude would be triggered when the cursor is moving at over 80% of the full speed towards the target. Otherwise, the vibration is not triggered. Again, the chosen threshold was informed by the feedback from our pilot studies.

10.3.5.5 *Combinations of Color and Vibration*

In addition to the individual visual and vibrotactile cues, three combinations of color and vibration were also tested. They were color combined with increasing vibration (C + IV), color combined with decreasing vibration (C + DV), and color combined with threshold vibration (C + TV).

10.3.6 *Hypotheses*

Our experiment was designed to evaluate the efficiency and experience of different visual and vibrotactile cues as guidance in data visualization tasks. Through it, we aim to gain a better understanding of how these cues work differently for target unknown and path unknown types of guidance. According to the research on the bandwidth of different sensory systems [194], eyes (visual feedback) have a higher bandwidth than skin (tactile feedback). Thus, we expected visual cues to outperform vibrotactile cues.

Hypothesis 1. *Color cue elicits better performance in time and accuracy than vibrotactile cues.*

Furthermore, we expect vibrotactile cues along with visual cues might outperform individual (singular) cues, as it utilizes two sensory channels of users simultaneously.

Hypothesis 2. *Combinations of both visual and vibrotactile cues elicit better performance in time and accuracy than visual or vibrotactile cues alone.*

10.3.7 *Procedure*

We established a set of protocols before we started the experiment. First, the participants were asked to fill out a demographic questionnaire regarding age, gender, handedness, the health of eyes, and experience with computer-based graphic user interfaces before they started the experiment. Before each task, we went through the task and the different types of visual and vibrotactile cues to make sure the participants understood the experiment clearly. Then they had a few minutes to familiarize themselves with the prototype. After each task, the participants were asked to fill in a short experience questionnaire and briefly discuss their overall experiences with different cues in that task. All participants were also provided the opportunity to take a break between the two tasks. At the end of the sessions, they were also asked to talk about how the cues worked differently in the two tasks.

10.3.7.1 *Tasks*

The experiment was done in a within-subjects design. Half of the participants started with the selection task, and the other half started with navigation. During each task, there were three rounds of tests that consisted of 21 different scenarios in each round. The order of the 21 scenarios was randomly generated for each round.

In the selection task, participants were asked to brush a certain number of points in a scatter plot. The number of points that the participants should select varied between 15, 60, and 240. When participants brushed through the points, seven different cues (visual cue, three vibrotactile cues, and three combinations of both visual and vibrotactile cues) were generated accordingly to guide their actions. As such, we had 21 (3 numbers $\times$ 7 cues) different combinations of the target number of points and cues. Both, the target number of points and the type of cues, were generated in a random order. Three rounds of these 21 scenarios – a total of 63 (21 scenarios $\times$ 3 rounds) trials – were performed by each participant.

In the navigation task, participants were asked to navigate through a scatter plot to find a specific point on the screen. The scatter plot was generated with a normal distribution function. The distance of the target point from the starting point could be 2400, 4000 or 6000 pixels, and the direction of the target point was generated randomly. Similar to the selection task, when participants panned to the target points, seven different cues (visual, three vibration, or three combinations of both visual and vibration) were generated accordingly to guide their action. As such, we had 21 (3 distances $\times$ 7 cues) different combinations of distances and cues. Both the distances and types of cues were generated in a random order. Three rounds of these 21 scenarios – a total of 63 (21 scenarios $\times$ 3 rounds) trials – were performed by each participant.

10.3.7.2 *Variables*

In this study, the independent variables are the type of cues and target number of points for the selection task or distance for the navigation task. However, the type of cues is the main focus of our study.

The dependent variables are as following: In the selection task, both time and accuracy were measured. Time was measured through the moment participants pushed down the mouse button to brush until the moment they released the button and finished brushing. Accuracy was measured by the deviation of the selected number of points from the target number of points. The time and accuracy were measured for the last attempt of brushing, if participants made several attempts in one trial. In the navigation task, only time was measured. Time was measured from the moment participants started to pan through the visualization until the moment they clicked on the target point.

10.3.7.3 *Questionnaire*

While the experiment would provide us with user performance data like time and accuracy, subjective evaluation can help us to interpret the performance data better. Thus, we decided to include a 7-point Likert scale in our study to investigate the experience of different cues in the tasks. A short question on their overall experience with different cues was included and participants were asked to rate it as "extremely bad" to "extremely good" on the scale of 1 to 7.

10.3.7.4 *Interview*

A semi-structured interview was also conducted to help us explain the results from the performance and questionnaire data [85]. Seven questions focusing on the users' subjective experiences with each cue were proposed during the internal discussions and tested out with two participants during the pilot study. The questions revolved around participants' overall experience as well as the comparison of different cues. Such as, "did you notice anything unexpected or interesting?", "which feedback stood out the most for you?", and "how do you feel the feedback work differently for you in the two tasks?".

10.4 RESULTS

The user study took around 45 minutes for each participant. Typical time spent was around 10 minutes for the selection task, and around 20 minutes for the navigation task. Among the 14 participants, 12 of them finished both tasks for 63 trials. Two participants finished all the 63 trials for the selection task, but only finished 42 trials for the navigation task due to fatigue. However, as they only skipped the last round of trials, they still went through the same number of trials for each type of cue and scenario. Therefore, this is not impacting the results they had obtained up to that point. In total, 882 trials of the selection task (14 participants $\times$ 3 target numbers $\times$ 7 cues $\times$ 3 rounds of trials) and 840 trials of the navigation task were completed.

The results of our user study were automatically captured through logged timestamps as well as positions of each click with the corresponding feedback cue and scenario. The deviation from target number of points was recorded additionally for the selection task. From these raw data, we computed task durations and average selection accuracies per trial.

Table 10.2: Mean and Standard Deviation for the Number of Errors and Completion Time in the Selection Task

	C	**IV**	**DV**	**TV**	**C + IV**	**C + DV**	**C + TV**
AVG Number of Errors	2.08	3.30	2.28	1.54	2.22	2.03	1.10
SD of Errors	*1.16*	*2.05*	*2.13*	*2.11*	*1.24*	*2.06*	*0.62*
AVG Completion Time	6.10s	8.92s	8.66s	8.76s	6.69s	6.76s	5.68s
SD of Completion Time	*3.46s*	*3.79s*	*4.62s*	*3.34s*	*4.23s*	*2.36s*	*1.55s*

Table 10.3: Mean and Standard Deviation for Completion Time in the Navigation Task

	C	**IV**	**DV**	**TV**	**C + IV**	**C + DV**	**C + TV**
AVG Completion Time	14.93s	13.68s	15.33s	16.88s	15.06s	16.16s	14.95s
SD of Completion Time	*5.76s*	*5.88s*	*4.06s*	*5.77s*	*4.43s*	*4.73s*	*4.02s*

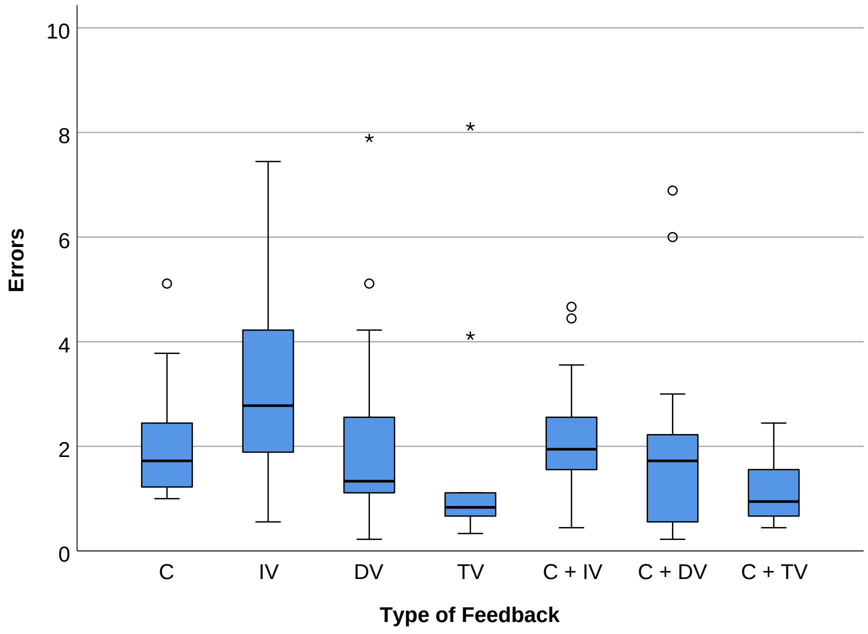


Figure 10.4: Boxplot of Errors in the Selection Task for Different Cues. Circles represent outliers (between 1.5 and 3 times the interquartile range) and asterisks represent the extremes (more than 3 times the interquartile range). The same applies to the following figures.

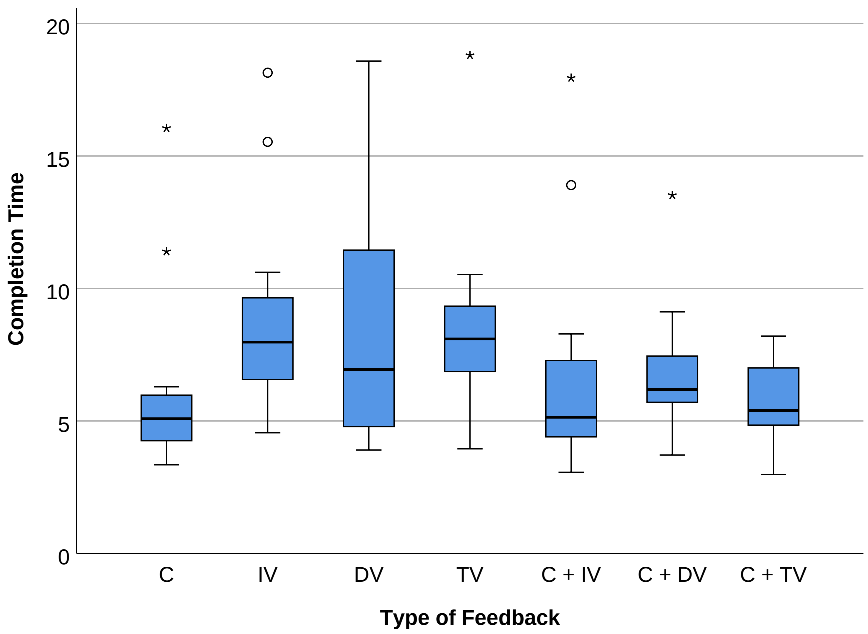


Figure 10.5: Boxplot of Completion Time in the Selection Task for Different Types of Feedback.

10.4.1 *Selection Task*

For the selection task, we calculated the means of performance time and accuracy for each participant in each type of cue and target number, then reported their mean and standard deviation.

10.4.1.1 *Time and Error*

For time, cues with only vibration performed the worst, while color with threshold vibration cue outperformed the color cue. For accuracy, color with threshold vibration cue had the highest accuracy, followed by threshold vibration individually. Accuracy for color, color with increasing vibration, color with decreasing vibration and decreasing vibration were similar to each other. Increasing vibration individually had the worst performance in both time and accuracy. Detailed results are presented in Table 10.2 and their boxplots are given in Figures 10.4 and 10.5.

As some of the results deviate from the normal distribution, a Friedman test adjusted by the Bonferroni correction for multiple tests was done to validate the difference between the results from different cues. For both error and time, the differences between the 7 cues are significant. However, the significance differs pairwise. For selection time, we summarize the following insights:

[ST1] C > {IV, TV} ($p < 0.01$)
Color had better (shorter) performance time than increasing vibration and threshold vibration.

[ST2] C + IV > IV ($p < 0.01$)
Color combined with increasing vibration had better (shorter) performance time than increasing vibration.

[ST3] C + TV > IV ($p < 0.01$)
Color combined with threshold vibration had better (shorter) performance time than increasing vibration.

For selection accuracy, we summarize the following insights:

[SA1] C + TV > IV ($p < 0.01$)
Color combined with threshold vibration was better (less error-prone) than increasing vibration.

[SA2] TV > IV ($p < 0.05$)
Threshold vibration was better (less error-prone) than increasing vibration.

[SA3] C + DV > IV ($p < 0.05$)
Color combined with decreasing vibration was better (less error-prone) than increasing vibration.

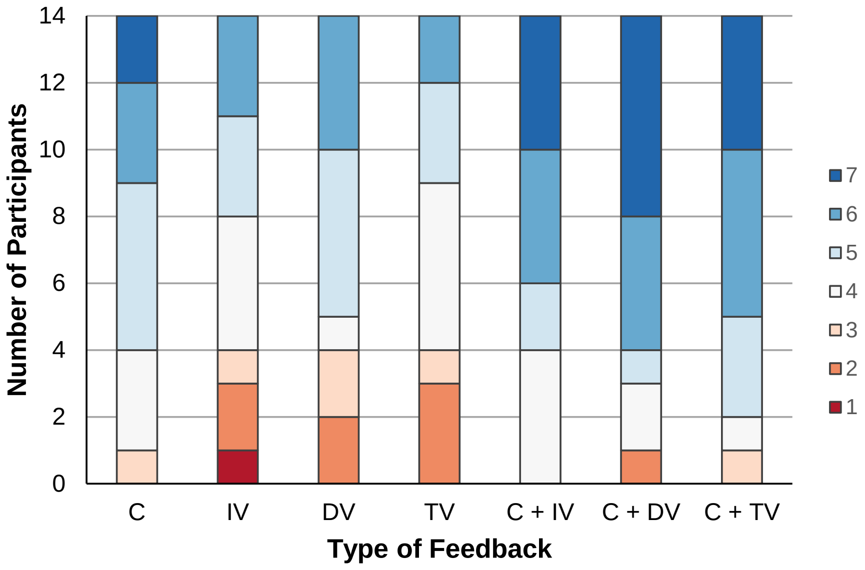


Figure 10.6: Distribution of Subjective Ratings in the Selection Task for Different Types of Feedback.

For the three tested scenarios (15, 60, and 240 targets), the time and number of errors both increased with more targets for each type of cue. No particular irregularities were observed from the results. Moreover, no significant learning effect was observed.

10.4.1.2 *Questionnaires*

A Friedman test adjusted by the Bonferroni correction for multiple tests was also done on the results from the questionnaire. The significant insights are summarized as following:

[SQ1] C + DV > {IV, TV} ($p < 0.05$)
Decreasing vibration with color was considered better than increasing vibration or threshold vibration individually.

The distribution of the questionnaire results is presented in Figure 10.6.

10.4.1.3 *Interviews*

These results were also reflected and further explained in our post-study interviews. Three participants (P1, P4, and P13) mentioned that the threshold vibration provided more sense of accuracy. Adding color cues to it helped them to find the rough area of the right number of points, while the threshold vibration allowed them to pinpoint the exact number of points to select. This gave them "a sense of security" (P1). Finally, for the increasing and decreasing vibration cues, some participants commented that decreasing vibration was better as "it is hard to tell when it is the highest vibration, but you know it when it stops vibrating" (P6).

10.4.1.4 *Summary*

Results from the selection task were mostly consistent with our hypotheses – overall color outperformed vibrotactile cues [ST1], while the appropriate combination of visual and vibrotactile cues, in this case, color and threshold vibration, outperformed color alone in accuracy ($p = 0.09$). Moreover, the performance in all combinations of visual and vibrotactile cues were improved compared to individual vibrotactile cues, although the significance varies [ST2, ST3, SA1, SA3, SQ1].

10.4.2 *Navigation Task*

For the navigation task, we calculated the means of performance time for each participant in each type of cue and distance, then reported their mean and standard deviation.

10.4.2.1 *Time*

Among the 7 different cues, increasing vibration individually performed the best, followed by color cue. For the combinations of color and vibration cues, results for both color with increasing vibration and with decreasing vibration was worse than their corresponding individual vibration cues, but color with threshold vibration outperformed the corresponding individual vibration cue. Detailed results are presented in Table 10.3. Their boxplot is shown in Figure 10.7. As done for the selection task, a Friedman test was also done for the performance in the navigation task. The only significant result before correction for multiple tests is:

[NT1] IV > TV ($p < 0.05$**)**
Users performed faster with increasing vibration than with threshold vibration.

However, there is no significant result after the Bonferroni correction for multiple tests.

For the three tested scenarios (2400, 4000 and 6000 pixels from the starting point), overall the completion time also increased with the target point further away from the starting point. Each participant went through three rounds of tests, and the average completion time was generally shorter in the later rounds.

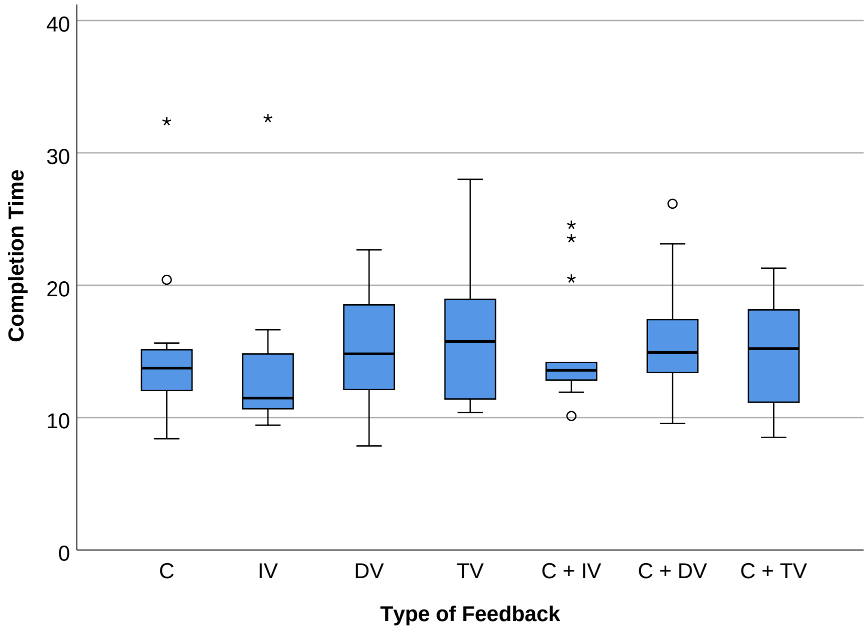


Figure 10.7: Boxplot of Completion Time in the Navigation Task for Different Types of Feedback.

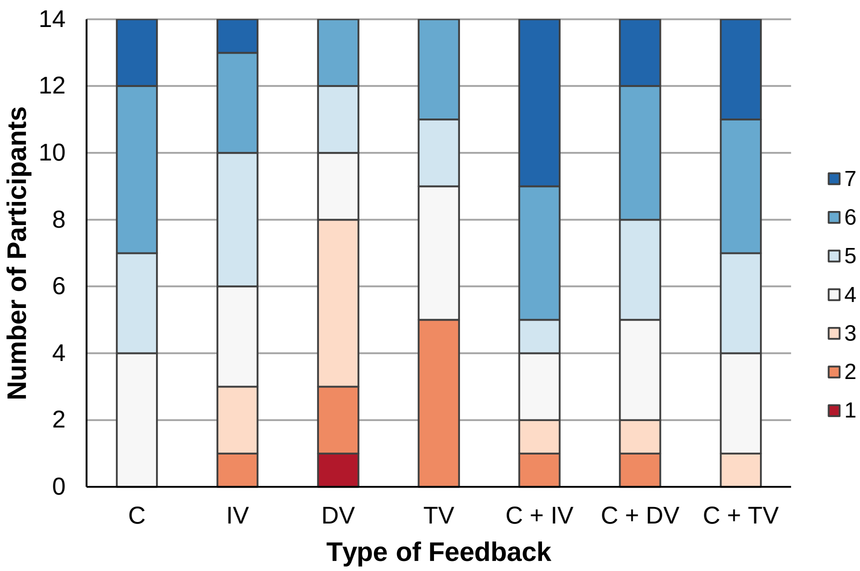


Figure 10.8: Distribution of Subjective Ratings in the Navigation Task for Different Types of Feedback.

10.4.2.2 *Questionnaires*

A Friedman test adjusted by the Bonferroni correction for multiple tests was also done on the results of the questionnaire:

[NQ1] {C + IV, C + TV} > DV ($p < 0.05$)
Increasing vibration with color and threshold vibration with color cues were considered better than decreasing vibration.

The distribution of the questionnaire results is presented in Figure 10.8.

10.4.2.3 *Interviews*

There were also some interesting insights from the post-study interviews of the navigation task. First, although most participants preferred the combinations of color and vibration cues, three participants (P2, P8, and P12) mentioned that the individual vibration cues were better than the combinations. In particular, one participant (P2) said that "you will worry more when the color is getting less dark, then you will panic and start to change the direction", but with the vibration "you will know right away if you are on the right direction". Second, the decreasing vibration was heavily criticized by several participants because "it vibrates all the time". However, on the metaphors of different vibration, one participant (P14) mentioned that the decreasing vibration might be helpful. "The vibration is constant, and you are searching for 'nothing'. It feels more game-like for me." They further explained that the action of "searching for nothing" means they were looking for the direction of "no vibration", which made them more at ease when they were on the right track. This participant subsequently argued that such metaphor feels more consistent, as one is rewarded with a more relaxed, calm feedback when something is done correctly.

Finally, for the threshold vibration cue, several participants mentioned that it did not work for them because it is too hard to trigger, and they had to search for it for a long time, while adding the color cue to it helped to find the rough direction first (P1, P9, and P12).

10.4.2.4 *Summary*

In the navigation task, the results were mostly inconsistent with our hypotheses. Among vibration-only cues, only the individual increasing vibration cue had slightly better performance than the color cue. For both increasing and decreasing vibration cues, the performance deteriorated when they were combined with color cues, while the performance of threshold vibration was improved combined with color. However, these differences were not statistically significant, which might be due to the relatively small sample size and overall long performance time.

10.5 DISCUSSION

From our results, we conclude some meaningful insights for using vibrotactile guidance in data visualization.

First, the same cues can work differently under different guidance scenarios. For the selection task under the *target unknown* scenario, the threshold vibration cue facilitated better performance, higher accuracy in particular [SA2], than increasing vibration, while for the navigation task under the *path unknown* scenario, the threshold vibration cue had the worst performance among all cues, and is particularly worse than increasing vibration [NT1]. The results from the post-study interview suggested that it might be due to the fact that the threshold vibration cue is better suited to communicate discrete guidance such as the ideal number of selected points and unable to effectively represent continuous information like path and speed.

Second, the visual cue alone in some cases significantly outperformed [ST1] vibrotactile cues, while a combination of visual and vibrotactile cues for guidance might not necessarily improve the user performance compared to visual or vibrotactile cues alone, potentially under more mentally taxing tasks. In the selection tasks, both time and accuracy were improved when vibration cues were combined with color, especially the performance time for increasing vibration [ST2]. However, in the navigation task, none of the user performances for combinations of color and vibrations was significantly better than their corresponding vibration cues alone or color cues, and some even deteriorated, although not significantly. Indeed, some participants reported the navigation task was "harder" than the selection task, and the combination of both cues can be confusing and distracting for some of the participants.

Finally, user experience and user performance might differ for the same guidance cues. specially in the navigation task, user performance was the best with increasing vibrotactile cues (see Table 10.3 and [NT1]). However, the user experience rating was not the highest for increasing vibration (see Figure 10.8). This suggests that user experience should be considered in addition to user performance when designing vibrotactile guidance.

10.6 LIMITATIONS AND FUTURE WORK

After discussing on the design of vibrotactile guidance and our experiment on vibrotactile cues, there is still room for improvements and extensions. Here we identified some possible areas of research moving forward.

First, contextual user studies will provide additional insights. The tested tasks in our study were isolated from the context of usage. The intention was to exclude any other variables and focus on the user performance in these tested tasks. However, implementing them in an existing visual analytics system could provide new understanding of how vibrotactile cues work and should be designed to work in combination with complex user interfaces and as part of full-fledged analytic workflows.

Second, more vibrotactile patterns in user guidance can be further explored. In our study, three different patterns were prioritized due to the user study design. With seven combinations of cues and two tasks, the user study already ran about 45 minutes for each participant. We are also aware that the hardware in our user study limited the options we have, and might have caused some bias in the results due to its limitation of five discrete amplitude levels. Therefore, in future work, vibrotactile cues with different parameters, hardware as well as additional visual cues can be further explored and compared to help us understand how each pattern of vibrotactile cue would match which scenario and task, as well as how they should be used in accordance with different visual cues.

Moreover, standards for subjective evaluation metrics of vibrotactile guidance can be further investigated. While user performance can be easily evaluated with time and error, we found that traditional subjective evaluation metrics like the User Experience Questionnaire (UEQ) or System Usability Scale (SUS) did not align well with the comparison of different cues for guidance in our pilot study. This might be due to the fact that these questionnaires are meant to evaluate the overall experience and usability of an entire system instead of individual cues. As a result, the characteristics used by them are rather hard to ascribe to a single cue like vibration – e.g., "organized vs. cluttered" or "friendly vs. unfriendly" from the UEQ [228], or the statement "I thought there was too much inconsistency in the system" from the SUS [6]. A deeper investigation on how vibrotactile guidance can be evaluated subjectively to gain more insights on their user experience and usability constitutes thus a formidable research challenge for future work.

10.7 CONCLUSIONS

To facilitate user interactions in data visualization, especially under visually overloaded scenarios, and provide a more immersive experience, non-visual guidance offers a largely unexplored design option. In this paper, we opened up the potential design space for one promising non-visual form of guidance – vibrotactile guidance – and reported on an experiment with different vibrotactile cues under two guidance scenarios. Our results suggest that while certain vibrotactile cues and their combination with visual cues outperformed visual cues alone, some other combinations can actually deteriorate user performance. Therefore, extra caution on aligning the cues with corresponding tasks and scenarios should be taken. Furthermore, a recurring theme from the observations and interview was the distraction that continuous vibrotactile feedback introduces, and how it might be stressing in combination with more mentally demanding tasks and confusing when used with visual cues. These observations suggest that vibrotactile cues are probably best suited to provide guidance at particular instances (e.g., a short pulse at the threshold) instead of using them over periods of time.

The experimental prototypes, anonymized data as well as devices considered in the vibrotactile design space are available at `https://vis-au.github.io/vibrotactile/`

ACKNOWLEDGMENTS

We thank Marius Hogräfer, Anke van Oosterhout, Ida Larsen-Ledet, as well as the anonymous reviewers for their feedback on earlier iterations of this work. We are also grateful to all the participants of the user study for their time and efforts.

11 PAPER E: SKETCHY RENDERING

Sketchy Rendering to Aid the Recollection of Regular Visualizations

Michael Reidun Engelbrecht Larsen, Wenkai Han, and Hans-Jörg Schulz

Department of Computer Science, Aarhus University, Denmark

ABSTRACT

Some visualizations have a more regular visual appearance than others. For example, while stream graphs or force-directed network layouts feature a unique, almost organic look&feel, matrices or unit treemaps can become rather bland, grid-like visualizations in which one data item is hard to tell apart from the next. In this paper, we investigate the use of sketchy rendering for such grid-like visualizations to give them a slightly more unique look&feel themselves. We evaluate our approach in a lab study ($N = 16$) where participants were asked to re-find a given grid cell in regular and sketchy grids. We find that users who make conscious use of the sketchy features can benefit from certain forms of sketchy rendering in terms of task completion times.

11.1 INTRODUCTION

In the context of visualization, sketchy rendering is a drawing style that mimics the imperfections of human-drawn sketches – e.g., a slightly shaking hand, varying pressure along strokes, or lines overshooting at connection points or corners. The sketchy rendering style is usually applied by perturbing contours of geometric shapes and applying hachures to fill them in a manner a human would [281]. More recently, extensions for 3D graphics [154] and the use of stippling [95] have been investigated. So far, sketchy rendering has been shown to work reasonably well for indicating uncertainties in data [21, 45], yet not so much for de-emphasizing imputed values [242].

We propose another use of sketchy rendering: enhancing visualizations that have a very regular appearance so as to improve the discernibility and memorability of their individual visual elements. Such regular visualizations can, for example, be matrices exhibiting a grid-like look and feel that makes it very hard to orient oneself on a large scale and equally challenging to discern one matrix cell from another on a detailed scale (cf. Fig. 11.1). But also unit-size treemaps are known to produce very regular, grid-like layouts [283]. The most common way to counter the uniform appearance of these visualizations is to use some form of coloring or shading of the individual elements – e.g., using a striped pattern [75] or a multi-hue color mapping [283].

In this work, we aim to achieve a similar effect through the use of sketchy line drawing. There are several reasons for choosing sketchy lines over fill colors for this task:

- Sketchy lines are most likely not used as a visual variable to encode data in the first place, as they are a rather poor channel that can only accurately convey between three and four different ordinal levels [21].
- Sketchy lines are first and foremost perceived as an element of visual style, not so much as carrying actual meaning [21], so that added sketchiness is less prone to be misinterpreted as a data characteristic.
- Sketchy lines perturb the gridded outline of the visualization layout, but leave the interior of the visual element available for showing additional data attributes.

To this end, we make a two-fold contribution: First, we propose an extended algorithm for sketchy line drawings that is highly parametrizable and which can thus be modulated and altered to generate grids with different regional appearances. And second, we evaluate the generated sketchy grids in a user study that looks at whether sketchy grids improve visualization users' ability to re-find a specific element in them.

11.2 SKETCHY RENDERING FOR REGULAR VISUALIZATIONS

The rendering of complex line geometries, strokes, and textures to generate a rich and natural visual appearance has seen continued interest over the years. Early works include multiresolution curves [82] for editing the appearance of complex lines, as well as using differently styled lines for creating depth cues similar to hand-drawn sketches [73]. In visualization, these techniques have been used for the illustrative rendering of anatomical geometry [227] and of flow field datasets [80].

In our work, we are mainly interested in creating a highly parametrizable sketchy line drawing, so that we can modulate its parameters differently in different regions of a visualization to create

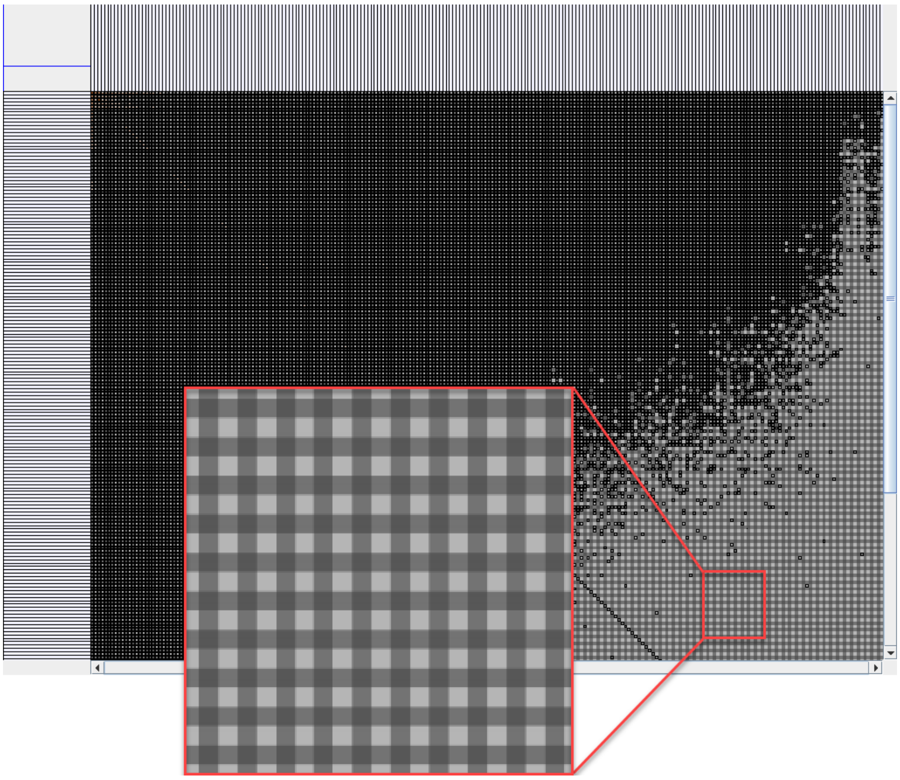


Figure 11.1: Adjacency matrix of Wikipedia data. Despite the striped pattern, the visualization exhibits in most parts a uniformity of display that impedes the recognition and recollection of data cells of interest. Image adapted from https://engineering.purdue.edu/~elm/projects/zame.html

unique regional appearances. We detail our approach in three steps: first we introduce our parametrizable sketchy line model, then we discuss its use for drawing rectangular shapes, and finally we show how to use either lines or shapes to generate entire layouts.

11.2.1 *Parametrizable Sketchy Lines*

Much of the literature is concerned with creating sketchy lines that are as close to hand-drawn lines as possible, going as far as using mathematical models of the physics behind human arm movements [1]. Our goal is different, as we are trying to generate as many distinctive looking line styles as possible without regarding their likeness to human drawings. Unhindered by the constraint of realism, we extend the line model introduced by Wood et al. [281]. In addition to the end points of a line, their model introduces two randomly placed control points that are then interpolated using Catmull-Rom-splines [35]. The intervals for placing the two control points are the only parameters of their model.

Our principal idea for extending this model is to increase the number of randomly placed control points, so that we have more flexibility and control over the appearance of each line. As a result, the following parameters govern the geometry of a line, with each of them also being illustrated in Figure 11.2:

- ***p*: The number of control points.** The more control points one uses, the more deviations from a straight line will appear, and thus the more ragged will the resulting sketchy line look.
- ***a*: The amount of displacement on each control point.** This amount captures how far from the straight line a control point can be placed – i.e., how much the line is contorted at each point. It is randomly chosen from an interval of $[-a, +a]$.
- ***d*: The distribution of the control points along the line.** Here we use the Bates distribution [9], where $d = 1$ yields a uniform distribution and $d \geqslant 2$ yields a probability distribution skewed towards the center of the line. For $d = -n$, we compute the "inverse" of the Bates distribution for n to generate a skew towards the endpoints of the line:

$$invBates(x, n) = \begin{cases} Bates(x + 0.5, n) & x < 0.5 \\ Bates(x - 0.5, n) & x \geqslant 0.5 \end{cases}$$

- ***o*: The order in which the control points are connected.** We specify the order as a list of offsets in which the points are used – e.g., $[+1]$ for a simple sequential order or $[+3, -1]$ to connect the points in order $1, 4, 3, 6, 5, 8, 7, 10$, etc. for an overdrawn line in a back&forth style.

Finally, we can also use the control points to vary the stroke width along the line through the following two parameters – both being illustrated in Figure 11.2(f):

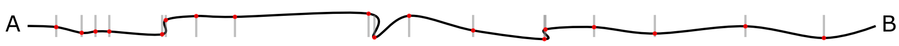
- ***sw*: The stroke width to be used.** If it shall be varied, this is usually an interval of permissible stroke widths from which a random value is chosen when changing the stroke.
- ***sc*: The count of points for which a stroke retains its width.** $sc = 1$ means that at every control point the stroke width will be changed, $sc = 2$ means that the stroke width is to be changed after every second point, and so forth until $sc = p$ means that the stroke width is only to be assigned once in the beginning.

11.2.2 *Turning Sketchy Lines into Shapes*

To draw shapes using our sketchy lines, we have in principle two fundamental options: concatenate the lines into a path or keep the

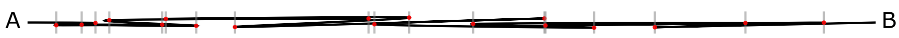
(a) A sketchy base line against which to compare the following.
$p = 18$, $a \leqslant 10$, $d = 1$, $o = [+1]$, $sw = 2$, $sc = 0$

(b) Increasing the number of control points.
$p = 36$, $a \leqslant 10$, $d = 1$, $o = [+1]$, $sw = 2$, $sc = 0$



(c) Increasing the amount of displacement at each control point.
$p = 18$, $a \leqslant 18$, $d = 1$, $o = [+1]$, $sw = 2$, $sc = 0$

(d) Placing the control points more towards the middle of the line.
$p = 18$, $a \leqslant 10$, $d = 3$, $o = [+1]$, $sw = 2$, $sc = 0$



(e) Jump over points in the connection order for back&forth effect.
$p = 18$, $a \leqslant 10$, $d = 1$, $o = [+3, -1]$, $sw = 2$, $sc = 0$

(f) Change stroke width at every third point to value between 2...8.
$p = 18$, $a \leqslant 10$, $d = 1$, $o = [+1]$, $sw \in [2..8]$, $sc = 3$

Figure 11.2: The different sketchy line parameters in effect.

lines disjoint. Both are illustrated in Figure 11.3. Concatenation allows us to generate shapes that look as if drawn with a single movement. This effect is reinforced by matching the stroke widths at the joints of the individual lines. The opposite effect is achieved by keeping the individual lines as such. Gaps and overspill at their connection points can be created by randomly displacing the endpoint of each line in a similar manner as described by Wood et al. [281].

Note that when rendering such sketchy shapes, a number of additional considerations need to be made, which include the style of the line cap and the need to combine the individual lines into a closed path if a fill color is to be assigned to the sketchy shape. With these aspects in place, any shape consisting of lines can be "sketchy-fied" using the extended sketchy line model from Section 11.2.1, as is demonstrated in Figure 11.4.

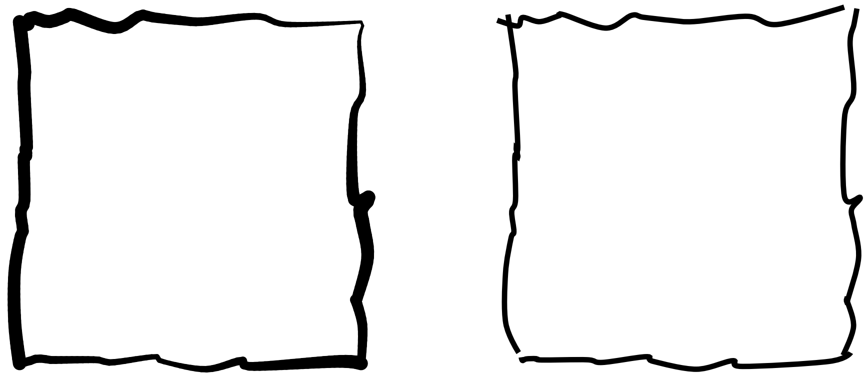


Figure 11.3: Different shape styles generated from the same four lines. Left: Concatenated lines to create a continuous shape. Right: Individual lines to create a shape as if drawn with multiple strokes.

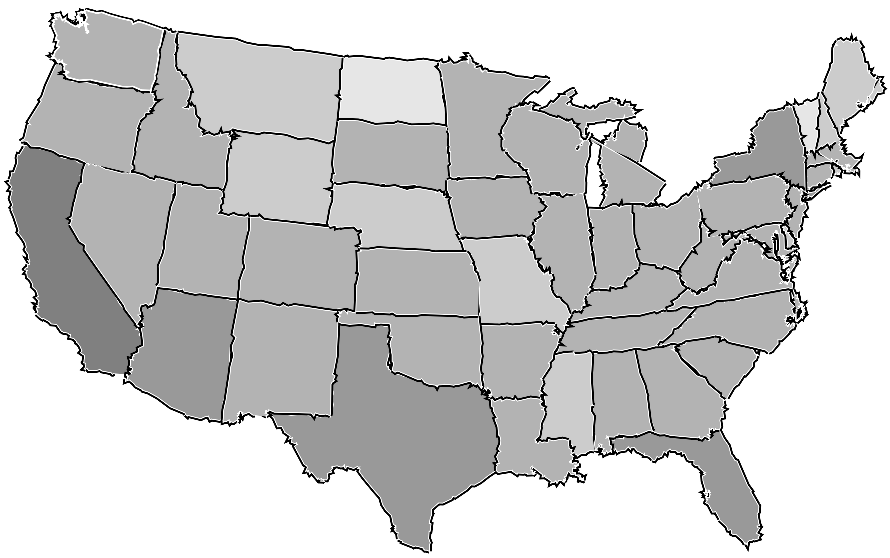


Figure 11.4: A sketchy map with each state/shape being "sketchy-fied" using the line parameters $p = 10$, $a = 4$, $d = 1$, $o = [+1]$, $sw = 5$, $sc = 0$. Image adapted from Wikipedia.

11.2.3 *Generating Sketchy Grid-Like Layouts*

For highly symmetric grid-like layouts, the literature is mainly concerned with means of distortion to either reflect properties of the data (e.g., registration errors [8]) or interactive changes (e.g., of the user focus [225]). In contrast to these approaches, we do not aim to distort the grid, just to perturb the grid line to some degree using our sketchy rendering.

For this, we propose two approaches: a *line-based approach* that pieces together the grid from individual sketchy lines of length equal to the grid cell size, and a *shape-based approach* that uses sketchy squares of grid cell size instead. By using these small lines/shapes, we can alter the appearance for each grid cell individually. This would

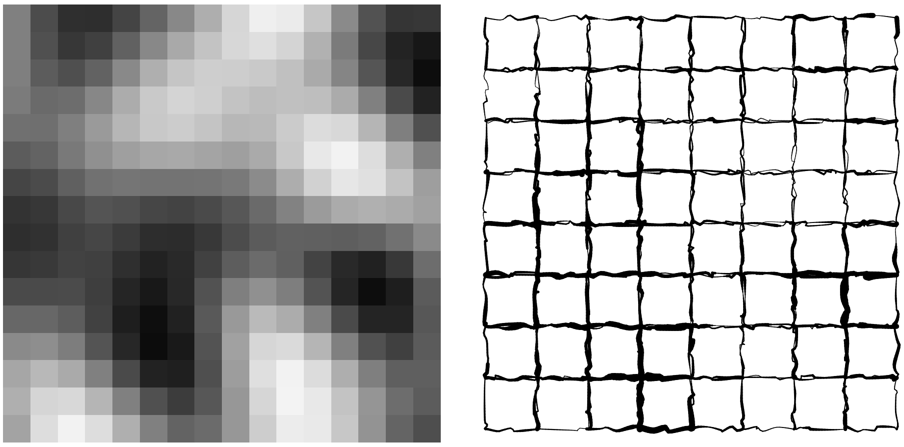


Figure 11.5: 2D Perlin noise field and its mapping onto $sw \in \{1..5\}$.

not be possible when, for example, using long sketchy lines that run across the whole grid.

To determine the line/shape parameters, we generate random 2D Perlin noise fields [204] – one for each of the four line parameters we vary: p, a, d, sw. An example of such a noise field can be seen on the left of Figure 11.5 with low values shown as white and high values shown as black. We then superimpose the grid on the noise fields. For the line-based approach, we determine the noise value at the middle of each individual line and parametrize it proportionally. This can be seen in the grid on the right side of Figure 11.5 that uses the noise field on the left to set the stroke widths sw of its sketchy lines. The shape-based approach does it likewise, but takes the noise value at the center of each grid cell and uses this value to proportionally parametrize all four sides of the shape in the same way. By using these random noise fields to assign parameter values instead of simply assigning any random parameter value, we yield parameter assignments that create smooth transitions without too abrupt changes between different styles of sketchy lines.

Setting the line parameters according to different Perlin noise fields yields grid regions with different line styles – from thin lines with barely noticeable jitter on both ends of the line to thick lines with a singular huge bump in the middle. The outcome for both grid styles – line-based and shape-based – can be seen in Figure 11.6.

11.3 Evaluation of Sketchiness for Recollecting Grid Locations

We created the sketchy grids with the intent of improving users' ability to orient themselves and to re-find cells of interest, as opposed to regular grids with straight lines. This section reports on a small lab study we did ($N = 16$, all students or researchers, ages between 23

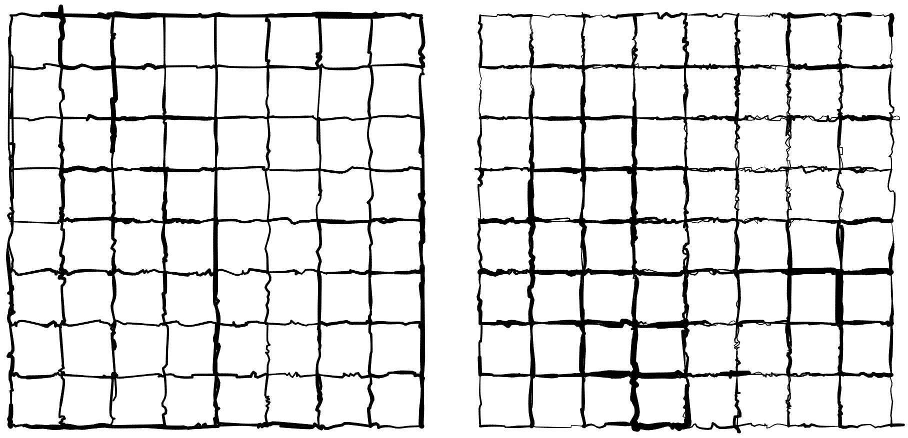


Figure 11.6: Line-based (left) and shape-based (right) grids with $p \in [1..18]$, $a \in [1..5]$, $d \in [\pm 1.. \pm 15]$, $o = [+1]$, $sw \in [1..5]$, $sc = 2$.

and 34, 5 female and 11 male) to evaluate whether users actually use the sketchy line features to this end and if so, whether it indeed leads to improvements.

11.3.1 *Setup of the Evaluation*

The participants were asked to perform a memorization task on 16×16 sized grids – straight-line grids, line-based sketchy grids, and shape-based sketchy grids, with the sketchy line parameters specified as in Figure 11.6. To randomize the order of the different grid types, the study was divided into 6 rounds. In each round, one trial was performed on each grid type with the order of the grid types being randomized. This results in $6 \times 3 = 18$ trials overall, which took the average participant around 10 minutes in total.

On these grids, for each trial one grid cell among the inner 14×14 cells was chosen at random, highlighted in red, and shown to the participant for 5 seconds. The restriction to inner grid cells was made to ensure that all highlighted cells are similar in the sense that they are surrounded by eight other cells. Afterwards, the participants were asked to answer a simple math question as a distraction, for which we used a 2-digit by 1-digit multiplication. Once completed, they were shown the same grid again and asked to re-find the location of the previously highlighted cell and click on it.

For each trial, we recorded the timestamp and position of every mouse click together with the position of the target cell and the grid type. From this raw data, we computed the following time and error metrics for further analysis:

- Time: For those trials in which the participants hit the target cell on their first try, this measure captures the time span between showing the grid and registering the mouse click.

- Error: For those trials in which the participants did not hit the target cell on their first try, this measure captures the Manhattan distance between the clicked cell and the target cell in the number of cells.

In case other metrics like the number of tries are to be used at a later point or to compare with other studies, we made the logged data available together with the software used in the evaluation at `https://vis-au.github.io/sketchyrendering`. Given this setup, our hypotheses were as follows:

H1: The average time for re-finding the highlighted cell on sketchy grids will be significantly lower than the one for straight-line grids.

H2: The average error of re-finding the highlighted cell on sketchy grids will be significantly lower than the one for straight line grids.

11.3.2 *Results of the Evaluation*

Among all participants, the results of the evaluation were inconclusive with the only statistically significant result ($p < 0.01$) being that line-based grids ($\mu = 4.135s$, $\sigma = 1.378s$) yielded better response times than the shape-based grids ($\mu = 5.297s$, $\sigma = 1.974s$).

Looking at the post-study interviews, it became clear that different participants followed different strategies in solving the task. This was to be expected, as we did not instruct the participants beforehand on the possible use of sketchiness and thus some discovered it as a feature that they actively and consciously used, while others did not pay much attention to it.

For the subset of those 6 participants who reported in the post-study interview to have used the sketchiness, we found that their response times with line-based sketchy grids ($\mu = 3.151s$, $\sigma = 0.729s$) were significantly better ($p < 0.05$) than with both – straight-line grids ($\mu = 5.237s$, $\sigma = 2.009s$) and shape-based grids ($\mu = 4.654s$, $\sigma = 1.548s$). Yet their differences in errors remained statistically inconclusive. The results for this subgroup are shown in Figure 11.7.

When asked in the interview what exactly their strategies for solving the task were, all participants reported to have counted in one way or another in most instances. The difference between the counting strategies was that the conscious users of the sketchiness counted from some nearby visual anchor produced by the sketchy rendering, whereas the other participants counted from the nearest borders or from the diagonals. This would explain the better response times with the line-based grids, as these produce more such visual anchors – e.g., tiny bends and wrinkles that serve as "landmarks" when trying to re-find a cell. The reason for this lies in the fact that shape-based grids overplot the lines between neighboring grid cells

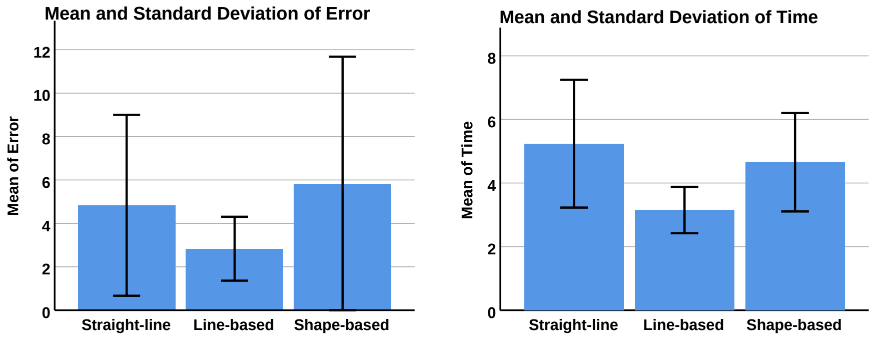


Figure 11.7: Means and standard deviations of error (left) and time (right) for participants who reported conscious use of sketchiness.

with the border lines of both adjacent squares. This makes it harder to use individual line features in shape-based grids, while such features appear clearly and without being overplotted in line-based grids. Thus, we can partially confirm H1 when users make conscious use of the sketchiness and uses a line-based sketchy grid.

11.4 conclusions

Sketchiness has been used in the past mainly to communicate levels of uncertainty. In this paper, we looked at a different possibility: using sketchy rendering to create more distinctive grid drawings. Through a small lab study, we gained partial confirmation for our intuition that the sketchy visual features could provide additional cues for orientating in very homogeneous visualizations.

We can easily envision this idea being applied beyond the realm of matrices and treemaps. For example, bubble charts with many bubbles being globbed together could benefit from a similar sketchy treatment – in particular, as their irregular circle-packing layouts lend themselves even less to counting strategies.

BIBLIOGRAPHY

[1] Z. AlMeraj, B. Wyvill, T. Isenberg, A. A. Gooch, and R. Guy. "Automatically Mimicking Unique Hand-Drawn Pencil Lines." In: *Computers & Graphics* 33.4 (2009), pp. 496–508. DOI: 10.1016/j.cag.2009.04.004.

[2] M. Angelini, R. Corriero, F. Franceschi, M. Geymonat, M. Mirabelli, C. Remondino, G. Santucci, and B. Stabellini. "A Visual Analytics System for Mobile Telecommunication Marketing Analysis." In: *Proceedings of the EuroVis Workshop on Visual Analytics*. The Eurographics Association, 2016, pp. 7–11. DOI: 10.2312/eurova.20161117.

[3] M. Angelini, T. May, G. Santucci, and H.-J. Schulz. "On Quality Indicators for Progressive Visual Analytics." In: *Proceedings of the EuroVis Workshop on Visual Analytics*. The Eurographics Association, 2019, pp. 25–29. DOI: 10.2312/eurova.20191120.

[4] M. Angelini, G. Santucci, H. Schumann, and H.-J. Schulz. "A Review and Characterization of Progressive Visual Analytics." In: *Informatics* 5.3 (2018), 31:1–27. DOI: 10.3390/informatics5030031.

[5] M. Azadi and L. A. Jones. "Evaluating Vibrotactile Dimensions for the Design of Tactons." In: *IEEE Transactions on Haptics* 7.1 (2014), pp. 14–23. DOI: 10.1109/TOH.2013.2296051.

[6] A. Bangor, P. T. Kortum, and J. T. Miller. "An Empirical Evaluation of the System Usability Scale." In: *International Journal of Human–Computer Interaction* 24.6 (2008), pp. 574–594. DOI: 10.1080/10447310802205776.

[7] G. Barbier and H. Liu. "Data Mining in Social Media." In: *Social Network Data Analytics*. Ed. by C. C. Aggarwal. Boston, MA: Springer US, 2011, pp. 327–352. ISBN: 978-1-4419-8462-3. DOI: 10.1007/978-1-4419-8462-3_12.

[8] L. Bastin, P. F. Fisher, and J. Wood. "Visualizing Uncertainty in Multi-Spectral Remotely Sensed Imagery." In: *Computers & Geosciences* 28.3 (2002), pp. 337–350. DOI: 10.1016/S0098-3004(01)00051-6.

[9] G. E. Bates. "Joint Distributions of Time Intervals for the Occurrence of Successive Accidents in a Generalized Polya Scheme." In: *The Annals of Mathematical Statistics* 26.4 (1955), pp. 705–720. DOI: 10.1214/aoms/1177728429.

[10] K. Bayoumy et al. "Smart Wearable Devices in Cardiovascular Care: Where We Are and How To Move Forward." In: *Nature Reviews Cardiology* 18.8 (2021), pp. 581–599. ISSN: 1759-5010. DOI: 10.1038/s41569-021-00522-7.

[11] E. Beauxis-Aussalet et al. "The Role of Interactive Visualization in Fostering Trust in AI." In: *IEEE Computer Graphics and Applications* 41.6 (2021), pp. 7–12. DOI: 10.1109/MCG.2021.3107875.

[12] M. Behrisch et al. "Quality Metrics for Information Visualization." In: *Computer Graphics Forum* 37.3 (2018), pp. 625–662. DOI: 10.1111/cgf.13446.

[13] V. Belton and T. Stewart. *Multiple Criteria Decision Analysis: An Integrated Approach*. Springer, 2002. DOI: 10.1007/978-1-4615-1495-4.

[14] S. Bensmaïa, M. Hollins, and J. Yau. "Vibrotactile Intensity and Frequency Information in the Pacinian System: A Psychophysical Model." In: *Perception & Psychophysics* 67.5 (2005), pp. 828–841. DOI: 10.3758/BF03193536.

[15] D. Bernoulli. "Exposition of a New Theory on the Measurement of Risk." In: *Econometrica* 22.1 (1954), pp. 23–36. ISSN: 00129682, 14680262. URL: http://www.jstor.org/stable/1909829.

[16] S. Bigaret, R. E. Hodgett, P. Meyer, T. Mironova, and A.-L. Olteanu. "Supporting the Multi-Criteria Decision Aiding Process: R and the MCDA Package." In: *EURO Journal on Decision Processes* 5.1-4 (2017), pp. 169–194. DOI: 10.1007/s40070-017-0064-1.

[17] D. Billman, G. Convertino, J. Shrager, P. Pirolli, and J. Massar. "Collaborative Intelligence Analysis With CACHE and Its Effects on Information Gathering and Cognitive Bias." In: *Human Computer Interaction Consortium Workshop*. 2006.

[18] F. Bolte and S. Bruckner. "Measures in Visualization Space." In: *Foundations of Data Visualization*. Ed. by M. Chen, H. Hauser, P. Rheingans, and G. Scheuermann. Springer, 2020, pp. 39–59. DOI: 10.1007/978-3-030-34444-3_3.

[19] 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*. Ed. by C. D. Hansen, M. Chen, C. R. Johnson, A. E. Kaufman, and H. Hagen. Springer, 2014, pp. 3–27. DOI: 10.1007/978-1-4471-6497-5_1.

[20] P. Booth, N. Gibbins, and S. Galanis. "Towards a Theory of Analytical Behaviour: A Model of Decision-Making in Visual Analytics." In: *Proceedings of the 52nd Hawaii International Conference on System Sciences*. This item is licensed under a Creative Commons Attribution Non-Commercial No Derivatives License. Honolulu: University of Hawaii at Manoa, 2019, pp. 1607–1616. URL: https://openaccess.city.ac.uk/id/eprint/21405/.

[21] N. Boukhelifa, A. Bezerianos, T. Isenberg, and J. Fekete. "Evaluating Sketchiness as a Visual Variable for the Depiction of Qualitative Uncertainty." In: *IEEE TVCG* 18.12 (2012), pp. 2769–2778. DOI: 10.1109/TVCG.2012.220.

[22] D. Bouyssou. "Outranking Methods." In: *Encyclopedia of Optimization*. Ed. by C. A. Floudas and P. M. Pardalos. Vol. 4. Springer, 2009, pp. 2887–2893. DOI: 10.1007/978-0-387-74759-0_495.

[23] M. Brehmer and T. Munzner. "A Multi-Level Typology of Abstract Visualization Tasks." In: *IEEE Transactions on Visualization and Computer Graphics* 19.12 (2013), pp. 2376–2385. DOI: 10.1109/TVCG.2013.124.

[24] N. Bressa, H. Korsgaard, A. Tabard, S. Houben, and J. Vermeulen. "What's the Situation with Situated Visualization? A Survey and Perspectives on Situatedness." In: *IEEE TVCG* 28.1 (2022), pp. 107–117. DOI: 10.1109/TVCG.2021.3114835.

[25] S. Brewster and L. M. Brown. "Tactons: Structured Tactile Messages for Non-Visual Information Display." In: *Proceedings of the Australasian User Interface Conference*. Australian Computer Society, 2004, pp. 15–23.

[26] K. Brodlie, R. A. Osorio, and A. Lopes. "A Review of Uncertainty in Data Visualization." In: *Expanding the Frontiers of Visual Analytics and Visualization*. Ed. by J. Dill, R. Earnshaw, D. Kasik, J. Vince, and P. C. Wong. Springer, 2012, pp. 81–109.

[27] E. T. Brown, A. Ottley, H. Zhao, Q. Lin, R. Souvenir, A. Endert, and R. Chang. "Finding Waldo: Learning About Users From Their Interactions." In: *IEEE TVCG* 20.12 (2014), pp. 1663–1672. DOI: 10.1109/TVCG.2014.2346575.

[28] L. M. Brown, S. A. Brewster, and H. C. Purchase. "A First Investigation Into the Effectiveness of Tactons." In: *Proceedings of the World Haptics Conference*. IEEE, 2005, pp. 167–176. DOI: 10.1109/WHC.2005.6.

[29] L. M. Brown, S. A. Brewster, and H. C. Purchase. "Multidimensional Tactons for Non-Visual Information Presentation in Mobile Devices." In: *Proceedings of the Conference on Human-*

Computer Interaction With Mobile Devices and Services. ACM, 2006, pp. 231–238. DOI: 10.1145/1152215.1152265.

[30] P. G. Brust-Renck, R. B. Weldon, and V. F. Reyna. "Judgment and Decision Making." In: *Oxford Research Encyclopedia of Psychology*. 2021. DOI: 10.1093/acrefore/9780190236557.013.536.

[31] F. Burstein and C. W. Holsapple. *Handbook on Decision Support Systems*. Springer, 2008. DOI: 10.1007/978-3-540-48713-5.

[32] V. L. Byrd and N. Dwenger. "Activity Worksheets for Teaching and Learning Data Visualization." In: *IEEE Computer Graphics and Applications* 41.6 (2021), pp. 25–36. DOI: 10.1109/MCG.2021.3115396.

[33] Á. A. Cabrera, W. Epperson, F. Hohman, M. Kahng, J. Morgenstern, and D. H. Chau. "FairVis: Visual Analytics for Discovering Intersectional Bias in Machine Learning." In: *Proceedings of IEEE VAST*. IEEE. 2019, pp. 46–56.

[34] C. Castelfranchi, R. Falcone, and E. Lorini. "A Non-Reductionist Approach to Trust." In: *Human-Computer Interaction*. 2009, pp. 45–72. ISBN: 9781848003552. DOI: 10.1007/978-1-84800-356-9.

[35] E. Catmull and R. Rom. "A Class of Local Interpolating Splines." In: *Computer Aided Geometric Design*. Ed. by R. E. Barnhill and R. F. Riesenfeld. Academic Press, 1974, pp. 317–326. DOI: 10.1016/B978-0-12-079050-0.50020-5.

[36] G. C. Cawley and N. L. Talbot. "On Over-Fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation." In: *The Journal of Machine Learning Research* 11 (2010), pp. 2079–2107. ISSN: 1532-4435.

[37] D. Ceneda, N. Andrienko, G. Andrienko, T. Gschwandtner, S. Miksch, N. Piccolotto, T. Schreck, M. Streit, J. Suschnigg, and C. Tominski. "Guide Me in Analysis: A Framework for Guidance Designers." In: *Computer Graphics Forum* 39.6 (2020), pp. 269–288. DOI: 10.1111/cgf.14017.

[38] D. Ceneda, T. Gschwandtner, T. May, S. Miksch, M. Streit, and C. Tominski. "Guidance or No Guidance? A Decision Tree Can Help." In: *Proceedings of the EuroVis Workshop on Visual Analytics*. The Eurographics Association, 2018, pp. 19–23. DOI: 10.2312/eurova.20181107.

[39] D. Ceneda, T. Gschwandtner, and S. Miksch. "A Review of Guidance Approaches in Visual Data Analysis: A Multifocal Perspective." In: *Computer Graphics Forum* 38.3 (2019), pp. 861–879. DOI: 10.1111/cgf.13730.

[40] D. Ceneda, T. Gschwandtner, and S. Miksch. "You Get by With a Little Help: The Effects of Variable Guidance Degrees on Performance and Mental State." In: *Visual Informatics* 3.4 (2019), pp. 177–191. DOI: 10.1016/j.visinf.2019.10.005.

[41] D. Ceneda, S. Miksch, T. Gschwandtner, H.-J. Schulz, M. Streit, T. May, and C. Tominski. "Characterizing Guidance in Visual Analytics." In: *IEEE TVCG* 23.1 (2017), pp. 111–120. DOI: 10.1109/TVCG.2016.2598468.

[42] 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." In: *Computer Graphics Forum* 39.3 (2020), pp. 713–756. DOI: 10.1111/cgf.14034.

[43] W.-R. Chen, Y.-H. Yun, M. Wen, H.-M. Lu, Z.-M. Zhang, and Y.-Z. Liang. "Representative Subset Selection and Outlier Detection via Isolation Forest." In: *Analytical Methods* 8.39 (2016), pp. 7225–7231. DOI: 10.1039/C6AY01574C.

[44] X. Chen, W. Zeng, Y. Lin, H. M. AI-maneea, J. Roberts, and R. Chang. "Composition and Configuration Patterns in Multiple-View Visualizations." In: *IEEE TVCG* 27.2 (2021), pp. 1514–1524. DOI: 10.1109/TVCG.2020.3030338.

[45] L. Cheong, C. Kinkeldey, I. Burfurd, S. Bleisch, and M. Duckham. "Evaluating the Impact of Visualization of Risk Upon Emergency Route-Planning." In: *International Journal of Geographical Information Science* (2019). to appear, pp. 1–29. DOI: 10.1080/13658816.2019.1701677.

[46] E. Chi, P. Barry, J. Riedl, and J. Konstan. "A Spreadsheet Approach to Information Visualization." In: *Proceedings of the IEEE Symposium on Information Visualization (InfoVis'97)*. IEEE, 1997, pp. 17–24. DOI: 10.1109/INFVIS.1997.636761.

[47] E. K. Chiou and J. D. Lee. "Trusting Automation: Designing for Responsivity and Resilience." In: *Human Factors* (2021). DOI: 10.1177/00187208211009995.

[48] I. Cho, R. Wesslen, A. Karduni, S. Santhanam, S. Shaikh, and W. Dou. "The Anchoring Effect in Decision-Making With Visual Analytics." In: *Proceedings of IEEE VAST*. IEEE. 2017, pp. 116–126.

[49] J. H. Cho, K. Chan, and S. Adali. "A Survey on Trust Modeling." In: *ACM Computing Surveys* 48.2 (2015), p. 28. ISSN: 15577341.

[50] L. Cibulski, E. Dimara, S. Hermawati, and J. Kohlhammer. "Supporting Domain Characterization in Visualization Design Studies With the Critical Decision Method." In: *Proceedings of the IEEE Workshop on Visualization Guidelines in Research, Design, and Education (VisGuides'22)*. to appear. IEEE, 2022.

[51] C. Collins, N. Andrienko, T. Schreck, J. Yang, J. Choo, U. Engelke, A. Jena, and T. Dwyer. "Guidance in the Human-Machine Analytics Process." In: *Visual Informatics* 2.3 (2018), pp. 166–180. DOI: 10.1016/j.visinf.2018.09.003.

[52] C. Conati, E. Hoque, D. Toker, and B. Steichen. "When to Adapt: Detecting User's Confusion During Visualization Processing." In: *Proceedings of the Workshop on User-Adaptive Visualization (WUAV'13)*. CEUR Workshop Proceedings, 2013, pp. 17–24.

[53] T. Connolly and B. K. Thorn. "Predecisional Information Acquisition: Effects of Task Variables on Suboptimal Search Strategies." In: *Organizational Behavior and Human Decision Processes* 39.3 (1987), pp. 397–416.

[54] K. Cook, N. Cramer, D. Israel, M. Wolverton, J. Bruce, R. Burtner, and A. Endert. "Mixed-Initiative Visual Analytics Using Task-Driven Recommendations." In: *Proceedings of IEEE VAST*. IEEE. 2015, pp. 9–16.

[55] E. Costante, J. Den Hartog, and M. Petkovic. "On-Line Trust Perception: What Really Matters." In: *Proceedings of the Workshop on Socio-Technical Aspects in Security and Trust (STAST)*. IEEE, 2011, pp. 52–59. ISBN: 9781457711817.

[56] J. C. Craig. "Vibrotactile Pattern Perception: Extraordinary Observers." In: *Science* 196.4288 (1977), pp. 450–452. DOI: 10.1126/science.850791.

[57] J. C. Craig and C. E. Sherrick. "Dynamic Tactile Displays." In: *Tactual Perception: A Sourcebook*. Ed. by W. Schiff and E. Foulke. Cambridge University Press, 1982, pp. 209–233.

[58] D. Danks and A. J. London. "Algorithmic Bias in Autonomous Systems." In: *Proceedings of IJCAI*. AAAI Press, 2017, pp. 4691–4697.

[59] A. Dasgupta, J.-Y. Lee, R. Wilson, R. A. Lafrance, N. Cramer, K. Cook, and S. Payne. "Familiarity vs Trust: A Comparative Study of Domain Scientists' Trust in Visual Analytics and Conventional Analysis Methods." In: *IEEE TVCG* 23.1 (2017), pp. 271–280. DOI: 10.1109/TVCG.2016.2598544.

[60] M. Daszykowski, B. Walczak, and D. Massart. "Representative Subset Selection." In: *Analytica Chimica Acta* 468.1 (2002), pp. 91–103. DOI: 10.1016/S0003-2670(02)00651-7.

[61] R. M. Dawes and B. Corrigan. "Linear Models in Decision Making." In: *Psychological Bulletin* 81.2 (1974), p. 95.

[62] C. Deck and S. Jahedi. "The Effect of Cognitive Load on Economic Decision Making: A Survey and New Experiments." In: *European Economic Review* 78 (2015), pp. 97–119. ISSN: 0014-2921. DOI: 10.1016/j.euroecorev.2015.05.004.

[63] M. Deller, A. Ebert, M. Bender, S. Agne, and H. Barthel. "Preattentive Visualization of Information Relevance." In: *Proceedings of the ACM International Multimedia Conference and Exhibition*. ACM, 2007, pp. 47–56. DOI: 10.1145/1290128.1290137.

[64] P. G. Devine. "Stereotypes and Prejudice: Their Automatic and Controlled Components." In: *Journal of Personality and Social Psychology* 56.1 (1989), p. 5.

[65] E. Dimara, G. Bailly, A. Bezerianos, and S. Franconeri. "Mitigating the Attraction Effect With Visualizations." In: *IEEE TVCG* 25.1 (2019), pp. 850–860. DOI: 10.1109/TVCG.2018.2865233.

[66] E. Dimara, S. Franconeri, C. Plaisant, A. Bezerianos, and P. Dragicevic. "A Task-Based Taxonomy of Cognitive Biases for Information Visualization." In: *IEEE TVCG* 26.2 (2018), pp. 1413–1432. DOI: 10.1109/TVCG.2018.2872577.

[67] E. Dimara and J. Stasko. "A Critical Reflection on Visualization Research: Where Do Decision Making Tasks Hide?" In: *IEEE TVCG* 28.1 (2021), pp. 1128–1138. DOI: 10.1109/TVCG.2021.3114813.

[68] B. Dougherty and A. Wade. *Vischeck Simulates Colourblind Vision*. (retrieved 23-AUG-2020). 2006. URL: http://www.vischeck.com/daltonize.

[69] P. Dourish. "What We Talk About When We Talk About Context." In: *Personal and Ubiquitous Computing* 8.1 (2004), pp. 19–30. DOI: 10.1007/s00779-003-0253-8.

[70] D. Dua and C. Graff. *UCI Machine Learning Repository*. (retrieved 01-NOV-2020). 2019. URL: http://archive.ics.uci.edu/ml.

[71] A. Edwards, G. Elwyn, J. Covey, E. Matthews, and R. Pill. "Presenting Risk Information a Review of the Effects of Framing and Other Manipulations on Patient Outcomes." In: *Journal of Health Communication* 6.1 (2001), pp. 61–82.

[72] H. J. Einhorn. "Expert Measurement and Mechanical Combination." In: *Organizational Behavior and Human Performance* 7.1 (1972), pp. 86–106.

[73] G. Elber. "Line Illustrations in Computer Graphics." In: *The Visual Computer* 11.6 (1995), pp. 290–296. DOI: 10.1007/BF01898406.

[74] G. Ellis. *Cognitive Biases in Visualizations*. Ed. by G. Ellis. Springer, 2018. ISBN: 9783319958309. DOI: 10.1007/978-3-319-95831-6.

[75] N. Elmqvist, T.-N. Do, H. Goodell, N. Henry, and J.-D. Fekete. "ZAME: Interactive Large-Scale Graph Visualization." In: *PacificVis'08: Proceedings of the IEEE Pacific Visualization Symposium*. Ed. by I. Fujishiro, H. Li, and K.-L. Ma. IEEE, 2008, pp. 215–222. DOI: 10.1109/PACIFICVIS.2008.4475479.

[76] T. Elrod, R. D. Johnson, and J. White. "A New Integrated Model of Noncompensatory and Compensatory Decision Strategies." In: *Organizational Behavior and Human Decision Processes* 95.1 (2004), pp. 1–19. ISSN: 0749-5978. DOI: 10.1016/j.obhdp.2004.06.002.

[77] M. J. Eppler and M. Aeschimann. "A Systematic Framework for Risk Visualization in Risk Management and Communication." In: *Risk Management* 11.2 (2009), pp. 67–89.

[78] J. M. Evangelista Belo, A. M. Feit, T. Feuchtner, and K. Grønbæk. "XRgonomics: Facilitating the Creation of Ergonomic 3D Interfaces." In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. CHI '21. Yokohama, Japan: ACM, 2021. ISBN: 9781450380966. DOI: 10.1145/3411764.3445349.

[79] J. S. B. Evans. "In Two Minds: Dual-Process Accounts of Reasoning." In: *Trends in Cognitive Sciences* 7.10 (2003), pp. 454–459.

[80] M. H. Everts, H. Bekker, J. B. T. M. Roerdink, and T. Isenberg. "Illustrative Line Styles for Flow Visualization." In: *Pacific Graphics'11: Short Paper Proceedings*. Ed. by B.-Y. Chen, J. Kautz, T.-Y. Lee, and M. C. Lin. The Eurographics Association, 2011, pp. 105–110. DOI: 10.2312/PE/PG/PG2011short/105-110.

[81] M. Figueres-Esteban, P. Hughes, and C. Van Gulijk. "The Role of Data Visualization in Railway Big Data Risk Analysis." In: *Proceedings of the 25th European Safety and Reliability Conference, ESREL 2015*. CRC Press/Balkema. 2015, pp. 2877–2882.

[82] A. Finkelstein and D. H. Salesin. "Multiresolution Curves." In: *SIGGRAPH'94: Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques*. Ed. by A. S. Glassner and M. Keeler. ACM, 1994, pp. 261–268. DOI: 10.1145/192161.192223.

[83] B. Friedman and H. Nissenbaum. "Bias in Computer Systems." In: *Acm Tois* 14.3 (1996), pp. 330–347.

[84] B. Frijns, A. Gilbert, T. Lehnert, and A. Tourani-Rad. "Uncertainty Avoidance, Risk Tolerance and Corporate Takeover Decisions." In: *Journal of Banking & Finance* 37.7 (2013), pp. 2457–2471.

[85] A. Galletta. *Mastering the Semi-Structured Interview and Beyond: From Research Design to Analysis and Publication*. NYU Press, 2013.

[86] W. Gani and M. Limam. "A Kernel Distance-Based Representative Subset Selection Method." In: *Journal of Statistical Computation and Simulation* 86.1 (2016), pp. 135–148. DOI: 10.1080/00949655.2014.996758.

[87] H. G. Gemünden and J. Hauschildt. "Number of Alternatives and Efficiency in Different Types of Top-Management Decisions." In: *European Journal of Operational Research* 22.2 (1985), pp. 178–190.

[88] J. F. George, K. Duffy, and M. Ahuja. "Countering the Anchoring and Adjustment Bias With Decision Support Systems." In: *Decision Support Systems* 29.2 (2000), pp. 195–206. DOI: 10.1016/S0167-9236(00)00074-9.

[89] G. Gigerenzer and D. G. Goldstein. "Reasoning the Fast and Frugal Way: Models of Bounded Rationality." In: *Psychological Review* 103.4 (1996), p. 650.

[90] J. Gill. *Vol. 2003 Royal National Institute of the Blind*. United Kingdom, 2003.

[91] L. Gitelman. *"Raw Data" Is an Oxymoron*. The MIT Press, Jan. 2013. ISBN: 9780262312325. DOI: 10.7551/mitpress/9302.001.0001.

[92] S. Gladisch, H. Schumann, and C. Tominski. "Navigation Recommendations for Exploring Hierarchical Graphs." In: *Proceedings of the International Symposium on Visual Computing (ISVC'13)*. Springer, 2013, pp. 36–47. DOI: 10.1007/978-3-642-41939-3_4.

[93] M. Gleicher, D. Albers, R. Walker, I. Jusufi, C. D. Hansen, and J. C. Roberts. "Visual Comparison for Information Visualization." In: *Information Visualization* 10.4 (2011), pp. 289–309. DOI: 10.1177/1473871611416549.

[94] J. Golbeck and U. Kuter. *Computing With Social Trust*. 2009, pp. 169–181. ISBN: 9781848003552. DOI: 10.1007/978-1-84800-356-9.

[95] J. Görtler, M. Spicker, C. Schulz, D. Weiskopf, and O. Deussen. "Stippling of 2D Scalar Fields." In: *IEEE TVCG* 25.6 (2019), pp. 2193–2204. DOI: 10.1109/TVCG.2019.2903945.

[96] D. Gotz, S. Sun, and N. Cao. "Adaptive Contextualization: Combating Bias During High-Dimensional Visualization and Data Selection." In: *Proceedings of the 21st International Conference on Intelligent User Interfaces*. ACM, 2016, pp. 85–95.

[97] D. Gotz and M. X. Zhou. "Characterizing Users' Visual Analytic Activity for Insight Provenance." In: *Information Visualization* 8.1 (2009), pp. 42–55. DOI: 10.1057/ivs.2008.31.

[98] E. Gunther. "Skinscape: A Tool for Composition in the Tactile Modality." PhD thesis. MIT, 2001.

[99] E. Gunther and S. O'Modhrain. "Cutaneous Grooves: Composing for the Sense of Touch." In: *International Journal of Phytoremediation* 32.4 (2003), pp. 369–381. DOI: 10.1076/jnmr.32.4.369.18856.

[100] D. Hägele, C. Schulz, C. Beschle, H. Booth, M. Butt, A. Barth, O. Deussen, and D. Weiskopf. "Uncertainty Visualization: Fundamentals and Recent Developments." In: *It – Information Technology* 64.4-5 (2022), pp. 121–132. DOI: 10.1515/itit-2022-0033.

[101] K. W. Hall, C. Perin, P. G. Kusalik, C. Gutwin, and S. Carpendale. "Formalizing Emphasis in Information Visualization." In: *Computer Graphics Forum* 35.3 (2016), pp. 717–737. DOI: 10.1111/cgf.12936.

[102] K. R. Hammond and T. R. Stewart. *The Essential Brunswik: Beginnings, Explications, Applications*. Oxford University Press, 2001.

[103] W. Han and H.-J. Schulz. "Beyond Trust Building — Calibrating Trust in Visual Analytics." In: *2020 IEEE Workshop on TRust and EXpertise in Visual Analytics (TREX)*. 2020, pp. 9–15.

[104] W. Han and H.-J. Schulz. "Exploring Vibrotactile Cues for Interactive Guidance in Data Visualization." In: *Proceedings of the 13th International Symposium on Visual Information Communication and Interaction*. ACM, 2020, 14:1–10. DOI: 10.1145/3430036.3430042.

[105] M. Harrower and C. A. Brewer. "ColorBrewer. Org: An Online Tool for Selecting Colour Schemes for Maps." In: *The Cartographic Journal* 40.1 (2003), pp. 27–37.

[106] M. G. Haselton, D. Nettle, and P. W. Andrews. "The Evolution of Cognitive Bias." In: *The Handbook of Evolutionary Psychology*. John Wiley & Sons, Ltd, 2015. Chap. 25, pp. 724–746. ISBN: 9780470939376. DOI: 10.1002/9780470939376.ch25.

[107] C. Heath and A. Tversky. "Preference and Belief: Ambiguity and Competence in Choice Under Uncertainty." In: *Journal of Risk and Uncertainty* 4.1 (1991), pp. 5–28.

[108] J. Heer and M. Agrawala. "Design Considerations for Collaborative Visual Analytics." In: *Information Visualization* 7.1 (2008), pp. 49–62.

[109] J. Heer and B. Shneiderman. "Interactive Dynamics for Visual Analysis." In: *Communications of the ACM* 55.4 (2012), pp. 45–54. DOI: 10.1145/2133806.2133821.

[110] M. Herschel, R. Diestelkämper, and H. B. Lahmar. "A Survey on Provenance: What For? What Form? What From?" In: *The VLDB Journal* 26.6 (2017), pp. 881–906.

[111] R. J. Heuer Jr. "Analysis of Competing Hypotheses." In: *Psychology of Intelligence Analysis*. US Govt. Printing Office, 1999, pp. 95–110.

[112] K. A. Hoff and M. Bashir. "Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust." In: *Human Factors* 57.3 (2015), pp. 407–434. ISSN: 15478181.

[113] H. Hofmann and M. Vendettuoli. "Common Angle Plots as Perception-True Visualizations of Categorical Associations." In: *IEEE TVCG* 19.12 (2013), pp. 2297–2305. DOI: 10.1109/TVCG.2013.140.

[114] D. Holliday, S. Wilson, and S. Stumpf. "User Trust in Intelligent Systems: A Journey Over Time." In: *Proceedings of the 21st International Conference on Intelligent User Interfaces*. IUI '16. Sonoma, California, USA: ACM, 2016, pp. 164–168. ISBN: 9781450341370. DOI: 10.1145/2856767.2856811.

[115] K. J. Holyoak and R. G. Morrison. *The Cambridge Handbook of Thinking and Reasoning*. Cambridge University Press, 2005.

[116] K. Höök and J. Löwgren. "Strong Concepts: Intermediate-Level Knowledge in Interaction Design Research." In: *ACM Trans. Comput.-Hum. Interact.* 19.3 (Oct. 2012). ISSN: 1073-0516. DOI: 10.1145/2362364.2362371.

[117] F. Huber, S. Köcher, J. Vogel, and F. Meyer. "Dazing Diversity: Investigating the Determinants and Consequences of Decision Paralysis." In: *Psychology & Marketing* 29.6 (2012), pp. 467–478. DOI: 10.1002/mar.20535.

[118] J. Hullman, X. Qiao, M. Correll, A. Kale, and M. Kay. "In Pursuit of Error: A Survey of Uncertainty Visualization Evaluation." In: *IEEE TVCG* 25.1 (2019), pp. 903–913. DOI: 10.1109/TVCG.2018.2864889.

[119] F. Hutter, L. Xu, H. H. Hoos, and K. Leyton-Brown. "Algorithm Runtime Prediction: Methods & Evaluation." In: *Artificial Intelligence* 206 (2014), pp. 79–111. DOI: 10.1016/j.artint.2013.10.003.

[120] A. Inc. *Expanding the Sensory Experience With Core Haptics - WWDC 2019 - Videos*. (retrieved 01-NOV-2020). 2019. URL: https://developer.apple.com/videos/play/wwdc2019/223/.

[121] A. Inc. *CHHapticEvent | Apple Developer Documentation*. (retrieved 01-NOV-2020). 2020. URL: https://developer.apple.com/documentation/corehaptics/chhapticevent.

[122] P. Isenberg, T. Isenberg, T. Hesselmann, B. Lee, U. von Zadow, and A. Tang. "Data Visualization on Interactive Surfaces: A Research Agenda." In: *IEEE Computer Graphics and Applications* 33.2 (2013), pp. 16–24. DOI: 10.1109/MCG.2013.24.

[123] H. Jänicke and G. Scheuermann. "Steady Visualization of the Dynamics in Fluids Using Varepsilon-Machines." In: *Computers & Graphics* 33.5 (2009), pp. 597–606.

[124] Y. Jansen, P. Dragicevic, P. Isenberg, J. Alexander, A. Karnik, J. Kildal, S. Subramanian, and K. Hornbæk. "Opportunities and Challenges for Data Physicalization." In: *Proceedings of the ACM Conference on Human Factors in Computing Systems*. ACM, 2015, pp. 3227–3236. DOI: 10.1145/2702123.2702180.

[125] N. Japkowicz and M. Shah. *Evaluating Learning Algorithms: A Classification Perspective*. Cambridge University Press, 2011. DOI: 10.1017/CBO9780511921803.

[126] W. Javed and N. Elmqvist. "Exploring the Design Space of Composite Visualization." In: *Proceedings of the IEEE Pacific Visualization Symposium (PacificVis'12)*. IEEE, 2012, pp. 1–8. DOI: 10.1109/PacificVis.2012.6183556.

[127] W. Jeong. "Multimodal Trivariate Thematic Maps With Auditory and Haptic Display." In: *Proceedings of the American Society for Information Science and Technology* 42.1 (2006). DOI: 10.1002/meet.14504201105.

[128] W. Jeong and M. Gluck. "Multimodal Bivariate Thematic Maps: Auditory and Haptic Display." In: *Proceedings of the American Society for Information Science and Technology* 39.1 (2002), pp. 279–283. DOI: 10.1002/meet.1450390130.

[129] W. Jeong and M. Gluck. "Multimodal Geographic Information Systems: Adding Haptic and Auditory Display." In: *Journal of the American Society for Information Science and Technology* 54.3 (2003), pp. 229–242. DOI: 10.1002/asi.10202.

[130] J. Jo, S. L'Yi, B. Lee, and J. Seo. "ProReveal: Progressive Visual Analytics With Safeguards." In: *IEEE TVCG* 27.7 (2019), pp. 3109–3122. ISSN: 19410506. DOI: 10.1109/TVCG.2019.2962404.

[131] B. J. Kadlec, H. M. Tufo, and G. A. Dorn. "Knowledge-Assisted Visualization and Segmentation of Geologic Features." In: *IEEE Computer Graphics and Applications* 30.1 (2009), pp. 30–39.

[132] D. Kahneman and A. Tversky. "Prospect Theory: An Analysis of Decision Under Risk." In: *Econometrica* 47.2 (1979), pp. 263–292.

[133] H. G. Kaper, E. Wiebel, and S. Tipei. "Data Sonification and Sound Visualization." In: *Computing in Science Engineering* 1.4 (1999), pp. 48–58. DOI: 10.1109/5992.774840.

[134] D. Keim, G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon. "Visual Analytics: Definition, Process, and Challenges." In: *Information Visualization*. Ed. by A. Kerren, J. T. Stasko, J.-D. Fekete, and C. North. Springer, 2008, pp. 154–175.

[135] D. A. Keim, F. Mansmann, J. Schneidewind, J. Thomas, and H. Ziegler. "Visual Analytics: Scope and Challenges." In: *Visual Data Mining: Theory, Techniques and Tools for Visual Analytics*. Ed. by S. J. Simoff, M. H. Böhlen, and A. Mazeika. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 76–90. DOI: 10.1007/978-3-540-71080-6_6.

[136] K. Kelton, K. R. Fleischmann, and W. A. Wallace. "Trust in Digital Information." In: *Journal of the American Society for Information Science and Technology* 59.3 (2008), pp. 363–374.

[137] N. Kerracher and J. Kennedy. "Constructing and Evaluating Visualisation Task Classifications: Process and Considerations." In: *Computer Graphics Forum* 36.3 (2017), pp. 47–59. DOI: 10.1111/cgf.13167.

[138] M. E. Kite, B. E. Whitley, and L. S. Wagner. *Psychology of Prejudice and Discrimination*. Routledge, 2022. DOI: 10.4324/9780367809218.

[139] 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 Conference on Naturalistic Decision Making*. Lawrence Erlbaum Assoc. 2007, pp. 113–155.

[140] K. J. Kokjer. "The Information Capacity of the Human Fingertip." In: *IEEE Transactions on Systems, Man and Cybernetics* 17.1 (1987), pp. 100–102. DOI: 10.1109/TSMC.1987.289337.

[141] G. Kou, Y. Peng, and G. Wang. "Evaluation of Clustering Algorithms for Financial Risk Analysis Using MCDM Methods." In: *Information Sciences* 275 (2014), pp. 1–12. DOI: 10.1016/j.ins.2014.02.137.

[142] C. Kozyrkov. *Your Dataset Is a Giant Inkblot Test*. (retrieved 23-AUG-2020). 2017. URL: https://towardsdatascience.com/your-dataset-is-a-giant-inkblot-test-b9bf4c53eec5.

[143] Z. Kuc. "Bidirectional Vibrotactile Communication System: Tactual Display Design and Attainable Data Rates." In: *Proceedings of VLSI and Computer Peripherals*. IEEE, 1989, pp. 2/101–2/103. DOI: 10.1109/CMPEUR.1989.93383.

[144] M. Kurze. "TGuide: A Guidance System for Tactile Image Exploration." In: *Proceedings of the International ACM Conference on Assistive Technologies*. ACM, 1998, pp. 85–91. DOI: 10.1145/274497.274514.

[145] T. C. Kwok, P. Kiefer, and M. Raubal. "Two-Step Gaze Guidance." In: ICMI '22. Bengaluru, India: ACM, 2022, pp. 299–309. ISBN: 9781450393904. DOI: 10.1145/3536221.3556612.

[146] T. von Landesberger. "Insights by Visual Comparison: The State and Challenges." In: *IEEE Computer Graphics and Applications* 38.3 (2018), pp. 140–148. DOI: 10.1109/MCG.2018.032421661.

[147] T. von Landesberger, S. Fiebig, S. Bremm, A. Kuijper, and D. W. Fellner. "Interaction Taxonomy for Tracking of User Actions in Visual Analytics Applications." In: *Handbook of Human Centric Visualization*. Ed. by W. Huang. Springer, 2014, pp. 653–670. DOI: 10.1007/978-1-4614-7485-2_26.

[148] E. Larson and C. Gray. *Project Management: The Managerial Process*. 8th. McGraw Hill, 2020.

[149] G. Leduc et al. "Road Traffic Data: Collection Methods and Applications." In: *Working Papers on Energy, Transport and Climate Change* 1.55 (2008), pp. 1–55.

[150] J. D. Lee and K. A. See. "Trust in Automation: Designing for Appropriate Reliance." In: *Human Factors* 46.1 (2004), pp. 50–80.

[151] M. D. Lee and T. D. Cummins. "Evidence Accumulation in Decision Making: Unifying the "Take the Best" and the "Rational" Models." In: *Psychonomic Bulletin & Review* 11.2 (2004), pp. 343–352.

[152] R. A. Leite, A. Arleo, J. Sorger, T. Gschwandtner, and S. Miksch. "Hermes: Guidance-Enriched Visual Analytics for Economic Network Exploration." In: *Visual Informatics* 4.4 (2020), pp. 11–22. DOI: 10.1016/j.visinf.2020.09.006.

[153] R. J. Lempert and S. Turner. "Engaging Multiple Worldviews With Quantitative Decision Support: A Robust Decision-Making Demonstration Using the Lake Model." In: *Risk Analysis* 41.6 (2021), pp. 845–865. DOI: 10.1111/risa.13579.

[154] D. Limberger, C. Fiedler, S. Hahn, M. Trapp, and J. Döllner. "Evaluation of Sketchiness as a Visual Variable for 2.5D Treemaps." In: *IV'16: Proceedings of the 20th International Conference Information Visualisation*. Ed. by E. Banissi et al. IEEE, 2016, pp. 183–189. DOI: 10.1109/IV.2016.61.

[155] J. Liu, N. Boukhelifa, and J. R. Eagan. "Understanding the Role of Alternatives in Data Analysis Practices." In: *IEEE TVCG* 26.1 (2020), pp. 66–76. DOI: 10.1109/TVCG.2019.2934593.

[156] S. Liu, A. H. Duffy, R. I. Whitfield, and I. M. Boyle. "Integration of Decision Support Systems to Improve Decision Support Performance." In: *Knowledge and Information Systems* 22.3 (2010), pp. 261–286. DOI: 10.1007/s10115-009-0192-4.

[157] Y. Liu, T. Althoff, and J. Heer. "Paths Explored, Paths Omitted, Paths Obscured: Decision Points & Selective Reporting in End-to-End Data Analysis." In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, 2020, pp. 1–14. DOI: 10.1145/3313831.3376533.

[158] Y. Liu, A. Kale, T. Althoff, and J. Heer. "Boba: Authoring and Visualizing Multiverse Analyses." In: *IEEE TVCG* 27.2 (2020), pp. 1753–1763. DOI: 10.1109/TVCG.2020.3028985.

[159] C. G. Lord, M. R. Lepper, and E. Preston. "Considering the Opposite: A Corrective Strategy for Social Judgment." In: *Journal of Personality and Social Psychology* 47.6 (1984), p. 1231.

[160] M. Luboschik, C. Maus, H.-J. Schulz, H. Schumann, and A. M. Uhrmacher. "Heterogeneity-Based Guidance for Exploring Multiscale Data in Systems Biology." In: *Proceedings of BioVis*. Ed. by J. Roerdink and M. Hibbs. IEEE, 2012, pp. 33–40. DOI: 10.1109/BioVis.2012.6378590.

[161] J. S. P. Macdonald and N. Lavie. "Visual Perceptual Load Induces Inattentional Deafness." In: *Attention, Perception, & Psychophysics* 73.6 (2011), pp. 1780–1789. ISSN: 1943-393X. DOI: 10.3758/s13414-011-0144-4.

[162] J. Mackinlay, P. Hanrahan, and C. Stolte. "Show Me: Automatic Presentation for Visual Analysis." In: *IEEE TVCG* 13.6 (2007), pp. 1137–1144. DOI: 10.1109/TVCG.2007.70594.

[163] K. MacLean and M. Enriquez. "Perceptual Design of Haptic Icons." In: *Proceedings of EuroHaptics*. 2003, pp. 351–363.

[164] M. MAGUIRE. "Context of Use within Usability Activities." In: *International Journal of Human-Computer Studies* 55.4 (2001), pp. 453–483. ISSN: 1071-5819. DOI: https://doi.org/10.1006/ijhc.2001.0486.

[165] T. Mahatody, M. Sagar, and C. Kolski. "State of the Art on the Cognitive Walkthrough Method, Its Variants and Evolutions." In: *International Journal of Human-Computer Interaction* 26.8 (2010), pp. 741–785. DOI: 10.1080/10447311003781409.

[166] S. Makonin, D. McVeigh, W. Stuerzlinger, K. Tran, and F. Popowich. "Mixed-Initiative for Big Data: The Intersection of Human + Visual Analytics + Prediction." In: *Proceedings of HICSS*. IEEE. 2016, pp. 1427–1436.

[167] J. Marks et al. "Design Galleries: A General Approach to Setting Parameters for Computer Graphics and Animation." In: *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH'97)*. ACM/Addison-Wesley, 1997, pp. 389–400. DOI: 10.1145/258734.258887.

[168] S. Marsh and P. Briggs. "Examining Trust, Forgiveness and Regret as Computational Concepts." In: *Human-Computer Interaction*. 2009, pp. 9–43. ISBN: 9781848003552. DOI: 10.1007/978-1-84800-356-9.

[169] S. Marsh and M. R. Dibben. "Trust, Untrust, Distrust and Mistrust – An Exploration of the Dark(er) Side." In: *Trust Management: Proceedings of iTrust*. Springer, 2005, pp. 17–33.

[170] L. Martignon, U. Hoffrage, A. R. Group, et al. "Why Does One-Reason Decision Making Work." In: *Simple Heuristics That Make Us Smart* (1999), pp. 119–140.

[171] T. May, A. Bannach, J. Davey, T. Ruppert, and J. Kohlhammer. "Guiding Feature Subset Selection With an Interactive Visualization." In: *Proceedings of IEEE VAST*. IEEE, 2011, pp. 111–120. DOI: 10.1109/VAST.2011.6102448.

[172] E. Mayr, N. Hynek, S. Salisu, and F. Windhager. "Trust in Information Visualization." In: *EuroVis Workshop on Trustworthy Visualization (TrustVis)*. Ed. by R. Kosara, K. Lawonn, L. Linsen, and N. Smit. The Eurographics Association, 2019. ISBN: 978-3-03868-091-8. DOI: 10.2312/trvis.20191187.

[173] B. McGuinness and A. Leggatt. "Factors Influencing Information Trust and Distrust in a Sensemaking Task." In: *Proceedings of the International Command and Control Research and Technology Symposium*. 2006.

[174] S. McKenna, A. Lex, and M. Meyer. "Worksheets for Guiding Novices Through the Visualization Design Process." In: 2017. DOI: 10.48550/arXiv.1709.05723.

[175] S. McKenna, D. Mazur, J. Agutter, and M. Meyer. "Design Activity Framework for Visualization Design." In: *IEEE TVCG* 20.12 (2014), pp. 2191–2200. DOI: 10.1109/TVCG.2014.2346331.

[176] I. C. McManus, K. Woolf, D. Harrison, P. A. Tiffin, L. W. Paton, K. Y. F. Cheung, and D. T. Smith. "Predictive validity of A-level grades and teacher-predicted grades in UK medical school applicants: a retrospective analysis of administrative data in a time of COVID-19." In: *BMJ Open* 11.12 (2021). ISSN: 2044-6055. DOI: 10.1136/bmjopen-2020-047354.

[177] A. McNutt and G. Kindlmann. "Linting for Visualization: Towards a Practical Automated Visualization Guidance System." In: *Proceedings of the 2nd VisGuides Workshop*. 2018.

[178] A. McNutt, G. Kindlmann, and M. Correll. "Surfacing Visualization Mirages." In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI '20. Honolulu, HI, USA: ACM, 2020, pp. 1–16. ISBN: 9781450367080. DOI: 10.1145/3313831.3376420.

[179] L. Micallef, H.-J. Schulz, M. Angelini, M. Aupetit, R. Chang, J. Kohlhammer, A. Perer, and G. Santucci. "The Human User in Progressive Visual Analytics." In: *EuroVis 2019*. Ed. by J. Johansson, F. Sadlo, and G. E. Marai. The Eurographics Association, 2019. ISBN: 978-3-03868-090-1. DOI: 10.2312/evs.20191164.

[180] Microsoft. *SimpleHapticsController Class (Windows.Devices.Haptics) - Windows UWP Applications*. (retrieved 01-NOV-2020). 2020. URL: https://docs.microsoft.com/en-us/uwp/api/windows.devices.haptics.simplehapticscontroller.

[181] M. Migut and M. Worring. "Visual Exploration of Classification Models for Risk Assessment." In: *Proceedings of IEEE VAST*. 2010, pp. 11–18. DOI: 10.1109/VAST.2010.5652398.

[182] H. Montgomery and O. Svenson. "On Decision Rules and Information Processing Strategies for Choices Among Multiattribute Alternatives." In: *Scandinavian Journal of Psychology* 17.1 (1976), pp. 283–291.

[183] 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." In: *IEEE TVCG* 25.1 (2019), pp. 438–448. DOI: 10.1109/TVCG.2018.2865240.

[184] T. Mühlbacher, H. Piringer, S. Gratzl, M. Sedlmair, and M. Streit. "Opening the Black Box: Strategies for Increased User Involvement in Existing Algorithm Implementations." In: *IEEE TVCG* 20.12 (2014), pp. 1643–1652. DOI: 10.1109/TVCG.2014.2346578.

[185] B. M. Muir. "Trust Between Humans and Machines, and the Design of Decision Aids." In: *International Journal of Man-Machine Studies* 27.5-6 (1987), pp. 527–539.

[186] E. Müller, I. Assent, R. Krieger, T. Jansen, and T. Seidl. "Morpheus: Interactive Exploration of Subspace Clustering." In: *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'08)*. ACM, 2008, pp. 1089–1092. DOI: 10.1145/1401890.1402026.

[187] T. Munzner. "A Nested Model for Visualization Design and Validation." In: *IEEE TVCG* 15.6 (2009), pp. 921–928. DOI: 10.1109/TVCG.2009.111.

[188] S. Murphy and P. Dalton. "Out of Touch? Visual Load Induces Inattentional Numbness." In: *Journal of Experimental Psychology: Human Perception and Performance* 42.6 (2016), pp. 761–765. DOI: 10.1037/xhp0000218.

[189] T. Mussweiler, F. Strack, and T. Pfeiffer. "Overcoming the Inevitable Anchoring Effect: Considering the Opposite Compensates for Selective Accessibility." In: *Personality and Social Psychology Bulletin* 26.9 (2000), pp. 1142–1150.

[190] A. A. Nalcaci, D. Girgin, S. Balki, F. Talay, H. A. Boz, and S. Balcisoy. "Detection of Confirmation and Distinction Biases in Visual Analytics Systems." In: *EuroVis Workshop on Trustworthy Visualization (TrustVis)*. Ed. by R. Kosara, K. Lawonn, L. Linsen, and N. Smit. The Eurographics Association, 2019. ISBN: 978-3-03868-091-8. DOI: 10.2312/trvis.20191185.

[191] A. Narechania, A. Coscia, E. Wall, and A. Endert. "Lumos: Increasing Awareness of Analytic Behavior during Visual Data Analysis." In: *IEEE TVCG* 28.1 (2022), pp. 1009–1018. DOI: 10.1109/TVCG.2021.3114827.

[192] B. R. Newell, D. A. Lagnado, and D. R. Shanks. *Straight Choices: The Psychology of Decision Making*. Psychology Press, 2015.

[193] D. A. Norman. "Some Observations on Mental Models." In: *Human-Computer Interaction: A Multidisciplinary Approach*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1987, pp. 241–244. ISBN: 0934613249.

[194] T. Nørretranders. *The User Illusion: Cutting Consciousness Down to Size*. Viking, 1991.

[195] L. Nowell, E. Hetzler, and T. Tanasse. "Change Blindness in Information Visualization: A Case Study." In: *Proceedings of InfoVis*. IEEE, 2001, pp. 15–22.

[196] 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: *Proceedings of EISIC*. IEEE, 2016, pp. 148–151.

[197] J. Ooge and K. Verbert. "Trust in Prediction Models: a Mixed-Methods Pilot Study on the Impact of Domain Expertise." In: *2021 IEEE Workshop on TRust and EXpertise in Visual Analytics (TREX)*. 2021, pp. 8–13. DOI: 10.1109/TREX53765.2021.00007.

[198] J. Ooge and K. Verbert. "Visually Explaining Uncertain Price Predictions in Agrifood: A User-Centred Case-Study." In: *Agriculture* 12.7 (2022). ISSN: 2077-0472. DOI: 10.3390/agriculture12071024.

[199] A. Ottley. *Adaptive and Personalized Visualization*. Morgan and Claypool, 2020.

[200] L. M. Padilla, S. H. Creem-Regehr, M. Hegarty, and J. K. Stefanucci. "Decision Making With Visualizations: A Cognitive Framework Across Disciplines." In: *Cognitive Research: Principles and Implications* 3.1 (2018), pp. 1–25.

[201] C. D. Parks and R. Cowlin. "Group Discussion as Affected by Number of Alternatives and by a Time Limit." In: *Organizational Behavior and Human Decision Processes* 62.3 (1995), pp. 267–275. DOI: 10.1006/obhd.1995.1049.

[202] B. Patnaik, A. Batch, and N. Elmqvist. "Information Olfactation: Harnessing Scent to Convey Data." In: *IEEE TVCG* 25.1 (2019), pp. 726–736. DOI: 10.1109/TVCG.2018.2865237.

[203] I. Pérez-Messina, D. Ceneda, M. El-Assady, S. Miksch, and F. Sperrle. "A Typology of Guidance Tasks in Mixed-Initiative Visual Analytics Environments." In: *Computer Graphics Forum* 41.3 (2022), pp. 465–476. DOI: 10.1111/cgf.14555.

[204] K. Perlin. "An Image Synthesizer." In: *ACM SIGGRAPH Computer Graphics* 19.3 (1985), pp. 287–296. DOI: 10.1145/325165.325247.

[205] E. J. Phares. *Locus of Control in Personality*. General Learning Press, 1976.

[206] R. A. Pick. "Benefits of Decision Support Systems." In: *Handbook on Decision Support Systems 1*. Springer, 2008, pp. 719–730.

[207] D. J. Power. "Decision Support Systems: A Historical Overview." In: *Handbook on Decision Support Systems 1*. Springer, 2008, pp. 121–140.

[208] H. Qian, R. Kuber, and A. Sears. "Towards Identifying Distinguishable Tactons for Use With Mobile Devices." In: *Proceedings of the International ACM Conference on Computers and Accessibility*. ACM, 2009, pp. 257–258. DOI: 10.1145/1639642.1639703.

[209] R. Raghunathan and M. T. Pham. "All Negative Moods Are Not Equal: Motivational Influences of Anxiety and Sadness on Decision Making." In: *Organizational Behavior and Human Decision Processes* 79.1 (1999), pp. 56–77.

[210] B. I. Reiner and E. Krupinski. "The Insidious Problem of Fatigue in Medical Imaging Practice." In: *Journal of Digital Imaging* 25.1 (Feb. 2012), pp. 3–6. ISSN: 1618-727X. DOI: 10.1007/s10278-011-9436-4.

[211] M. Riabacke, M. Danielson, and L. Ekenberg. "State-of-the-Art Prescriptive Criteria Weight Elicitation." In: *Advances in Decision Sciences* 2012 (2012), p. 276584. DOI: 10.1155/2012/276584.

[212] K. M. Richardson, S. D. Fouquet, E. Kerns, and R. J. McCulloh. "Impact of Mobile Device-Based Clinical Decision Support Tool on Guideline Adherence and Mental Workload." In: *Academic Pediatrics* 19.7 (2019), pp. 828–834. DOI: 10.1016/j.acap.2019.03.001.

[213] J. C. Roberts. "State of the Art: Coordinated & Multiple Views in Exploratory Visualization." In: *Proceedings of the 5th International Conference on Coordinated and Multiple Views in Exploratory Visualization (CMV'07)*. IEEE, 2007, pp. 61–71. DOI: 10.1109/CMV.2007.20.

[214] J. C. Roberts, C. J. Headleand, and P. D. Ritsos. *Five Design-Sheets: Creative Design and Sketching for Computing and Visualisation*. Springer, 2017. DOI: 10.1007/978-3-319-55627-7.

[215] J. C. Roberts. "On Encouraging Multiple Views for Visualization." In: *Proceedings the Conference on Information Visualisation (IV'98)*. IEEE, 1998, pp. 8–14. DOI: 10.1109/IV.1998.694193.

[216] G. Robertson, D. Ebert, S. Eick, D. Keim, and K. Joy. "Scale and Complexity in Visual Analytics." In: *Information Visualization* 8.4 (2009), pp. 247–253. DOI: 10.1057/ivs.2009.23.

[217] M. M. Roghanizad and D. J. Neufeld. "Intuition, Risk, and the Formation of Online Trust." In: *Computers in Human Behavior* 50 (2015), pp. 489–498.

[218] R. Rouw and N. B. Root. "Distinct Colours in the 'Synaesthetic Colour Palette'." In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 374.1787 (2019), p. 20190028. DOI: 10.1098/rstb.2019.0028.

[219] J. Rovan and V. Hayward. "Typology of Tactile Sounds and Their Synthesis in Gesture-Driven Computer Music Performance." In: *Trends in Gestural Control of Music*. Ed. by M. M. Wanderley and M. Battier. Editions IRCAM, 2000, pp. 297–320.

[220] T. L. Saaty. "What Is the Analytic Hierarchy Process?" In: *Mathematical Models for Decision Support*. Springer, 1988, pp. 109–121.

[221] D. Sacha, H. Senaratne, B. C. Kwon, G. Ellis, and D. A. Keim. "The Role of Uncertainty, Awareness, and Trust in Visual Analytics." In: *IEEE TVCG* 22.1 (2016), pp. 240–249. ISSN: 10772626. DOI: 10.1109/TVCG.2015.2467591.

[222] D. Sacha, A. Stoffel, F. Stoffel, B. C. Kwon, G. Ellis, and D. A. Keim. "Knowledge Generation Model for Visual Analytics." In: *IEEE TVCG* 20.12 (2014), pp. 1604–1613. DOI: 10.1109/TVCG.2014.2346481.

[223] M. Saha et al. "Visualizing Urban Accessibility: Investigating Multi-Stakeholder Perspectives through a Map-Based Design Probe Study." In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. CHI '22. New Orleans, LA, USA: ACM, 2022. ISBN: 9781450391573. DOI: 10.1145/3491102.3517460.

[224] W. Sałabun, J. Watróbski, and A. Shekhovtsov. "Are MCDA Methods Benchmarkable? A Comparative Study of TOPSIS, VIKOR, COPRAS, and PROMETHEE II Methods." In: *Symmetry* 12.9 (2020), 1549:1–56. DOI: 10.3390/sym12091549.

[225] M. Sarkar, S. S. Snibbe, O. J. Tversky, and S. P. Reiss. "Stretching the Rubber Sheet: A Metaphor for Viewing Large Layouts on Small Screens." In: *UIST'93: Proceedings of the 6th Annual ACM Symposium on User Interface Software and Technology*. Ed. by S. E. Hudson, R. Pausch, B. T. V. Zanden, and J. D. Foley. ACM, 1993, pp. 81–91. DOI: 10.1145/168642.168650.

[226] B. Scheibehenne, B. Von Helversen, and Y. Shevchenko. "Change and Status Quo in Decisions With Defaults: The Effect of Incidental Emotions Depends on the Type of Default." In: *Judgment and Decision Making* 9.3 (2014), pp. 287–296.

[227] S. Schlechtweg, B. Schönwälder, L. Schumann, and T. Strothotte. "Surfaces to Lines: Rendering Rich Line Drawings." In: *WSCG'98: Proceedings of the 6th International Conference in Central Europe on Computer Graphics and Visualization*. Ed. by V. Skala. 1998, pp. 354–361. URL: http://wscg.zcu.cz/wscg1998/papers98/schlechtwegwscg98.ps.gz.

[228] M. Schrepp, A. Hinderks, and J. Thomaschewski. "Construction of a Benchmark for the User Experience Questionnaire (UEQ)." In: *International Journal of Interactive Multimedia and Artificial Intelligence* 4.4 (2017), pp. 40–44. DOI: 10.9781/ijimai.2017.445.

[229] H.-J. Schulz, T. Nocke, M. Heitzler, and H. Schumann. "A Design Space of Visualization Tasks." In: *IEEE TVCG* 19.12 (2013), pp. 2366–2375. DOI: 10.1109/TVCG.2013.120.

[230] H.-J. Schulz, T. Nocke, M. Heitzler, and H. Schumann. "A Systematic View on Data Descriptors for the Visual Analysis of Tabular Data." In: *Information Visualization* 16.3 (2017), pp. 232–256.

[231] H.-J. Schulz, M. Streit, T. May, and C. Tominski. "Towards a Characterization of Guidance in Visualization." In: *Poster at the IEEE InfoVis*. 2013.

[232] M. Schweinsberg et al. "Same Data, Different Conclusions: Radical Dispersion in Empirical Results When Independent Analysts Operationalize and Test the Same Hypothesis." In: *Organizational Behavior and Human Decision Processes* 165 (2021), pp. 228–249. ISSN: 0749-5978. DOI: 10.1016/j.obhdp.2021.02.003.

[233] scikit-learn. *Clustering*. (retrieved 23-AUG-2022). 2022. URL: https://scikit-learn.org/stable/modules/clustering.html.

[234] S. Silva, B. S. Santos, and J. Madeira. "Using Color in Visualization: A Survey." In: *Computers & Graphics* 35.2 (2011), pp. 320–333.

[235] H. A. Simon. "Rational Choice and the Structure of the Environment." In: *Psychological Review* 63.2 (1956), p. 129.

[236] H. A. Simon. *The New Science of Management Decision*. The Ford distinguished lectures. Harper & Brothers, 1960.

[237] Y. Siskos, E. Grigoroudis, and N. F. Matsatsinis. "UTA Methods." In: *Multiple Criteria Decision Analysis: State of the Art Surveys*. Ed. by J. Figueira, S. Greco, and M. Ehrogott. Springer, 2005, pp. 297–334. DOI: 10.1007/0-387-23081-5_8.

[238] G. Slanzi, J. A. Balazs, and J. D. Velásquez. "Combining Eye Tracking, Pupil Dilation and EEG Analysis for Predicting Web Users Click Intention." In: *Information Fusion* 35 (2017), pp. 51–57.

[239] P. Slovic, M. L. Finucane, E. Peters, and D. G. MacGregor. "Risk as Analysis and Risk as Feelings: Some Thoughts About Affect, Reason, Risk, and Rationality." In: *Risk Analysis: An International Journal* 24.2 (2004), pp. 311–322.

[240] S. L. Smith and J. N. Mosier. *Guidelines for Designing User Interface Software*. Citeseer, 1986.

[241] B. Solhaug, D. Elgesem, and K. Stølen. "Why Trust Is Not Proportional to Risk." In: *Proceedings of ARES*. IEEE, 2007, pp. 11–18. ISBN: 0769527752. DOI: 10.1109/ARES.2007.161.

[242] H. Song and D. A. Szafir. "Where's My Data? Evaluating Visualizations With Missing Data." In: *IEEE TVCG* 25.1 (2019), pp. 914–924. DOI: 10.1109/TVCG.2018.2864914.

[243] F. Sperrle, H. Schäfer, D. Keim, and M. El-Assady. "Learning Contextualized User Preferences for Co-Adaptive Guidance in Mixed-Initiative Topic Model Refinement." In: *Computer Graphics Forum* 40.3 (2021), pp. 215–226. DOI: 10.1111/cgf.14301.

[244] F. Sperrle, D. Ceneda, and M. El-Assady. “Lotse: A Practical Framework for Guidance in Visual Analytics.” In: *IEEE Transactions on Visualization and Computer Graphics* (2022), pp. 1–11. DOI: 10.1109/TVCG.2022.3209393.

[245] F. Sperrle, A. Jeitler, J. Bernard, D. Keim, and M. El-Assady. “Co-Adaptive Visual Data Analysis and Guidance Processes.” In: *Computers & Graphics* 100 (2021), pp. 93–105. DOI: 10.1016/j.cag.2021.06.016.

[246] F. Sperrle, A. V. Jeitler, J. Bernard, D. A. Keim, and M. El-Assady. “Learning and Teaching in Co-Adaptive Guidance for Mixed-Initiative Visual Analytics.” In: *Proceedings of the EuroVis Workshop on Visual Analytics*. Eurographics, 2020, pp. 61–65.

[247] D. Sprague and M. Tory. “Exploring How and Why People Use Visualizations in Casual Contexts: Modeling User Goals and Regulated Motivations.” In: *Information Visualization* 11.2 (2012), pp. 106–123.

[248] I. O. for Standardization. *ISO/IEC 25063:2014 – Systems and Software Engineering; Systems and Software Product Quality Requirements and Evaluation (SQuaRE); Common Industry Format (CIF) for Usability: Context of Use Description*. (retrieved 23-AUG-2022). 2014. URL: https://www.iso.org/obp/ui/#iso:std:iso-iec:25063.

[249] I. O. for Standardization. *ISO 9241-11:2018 – Ergonomics of Human-System Interaction; Part 11: Usability: Definitions and Concepts*. (retrieved 23-AUG-2022). 2018. URL: https://www.iso.org/obp/ui/#iso:std:iso:9241:-11.

[250] SteelSeries. *Gamesense SDK*. (retrieved 01-NOV-2020). 2019. URL: https://github.com/SteelSeries/gamesense-sdk/blob/master/doc/api/json-handlers-tactile.md#reference-sections---ti-predefined-vibrations.

[251] “Steering data quality with visual analytics: The complexity challenge.” In: *Visual Informatics* 2.4 (2018), pp. 191–197. ISSN: 2468-502X. DOI: 10.1016/j.visinf.2018.12.001.

[252] M. Steinberger, M. Waldner, M. Streit, A. Lex, and D. Schmalstieg. “Context-Preserving Visual Links.” In: *IEEE TVCG* 17.12 (2011), pp. 2249–2258. DOI: 10.1109/TVCG.2011.183.

[253] C. Stoiber, D. Ceneda, M. Wagner, V. Schetinger, T. Gschwandtner, M. Streit, S. Miksch, and W. Aigner. “Perspectives of Visualization Onboarding and Guidance in VA.” In: *Visual Informatics* 6.1 (2022), pp. 68–83. DOI: 10.1016/j.visinf.2022.02.005.

[254] H. v. Storch and F. W. Zwiers. *Statistical Analysis in Climate Research*. Cambridge University Press, 1999. DOI: 10.1017/cbo9780511612336.

[255] M. Streit, H.-J. Schulz, A. Lex, D. Schmalstieg, and H. Schumann. "Model-Driven Design for the Visual Analysis of Heterogeneous Data." In: *IEEE TVCG* 18.6 (2011), pp. 998–1010. DOI: 10.1109/TVCG.2011.108.

[256] A. Tang, P. McLachlan, K. Lowe, C. R. Saka, and K. MacLean. "Perceiving Ordinal Data Haptically Under Workload." In: *Proceedings of the 7th International Conference on Multimodal Interfaces*. ACM, 2005, pp. 317–324. DOI: 10.1145/1088463.1088517.

[257] J. J. Thomas and K. A. Cook. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE Computer Society, 2005.

[258] C. Tominski and H. Schumann. *Interactive Visual Data Analysis*. AK Peters Visualization Series. CRC Press, 2020. ISBN: 9781498753999.

[259] C. Tominski, J. Abello, and H. Schumann. "CGV – An Interactive Graph Visualization System." In: *Computers and Graphics* 33.6 (2009), pp. 660–678. DOI: 10.1016/j.cag.2009.06.002.

[260] M. B. Tools. *Custom Notification Patterns*. (retrieved 01-NOV-2020). 2020. URL: https://help.mibandtools.com/knowledge_base/topics/custom-notification-patterns.

[261] M. Tory and T. Möller. "Evaluating Visualizations: Do Expert Reviews Work?" In: *IEEE Computer Graphics and Applications* 25.5 (2005), pp. 8–11. DOI: 10.1109/MCG.2005.102.

[262] B. M. Turner and D. R. Schley. "The Anchor Integration Model: A Descriptive Model of Anchoring Effects." In: *Cognitive Psychology* 90 (2016), pp. 1–47.

[263] A. Tversky and D. Kahneman. "The Framing of Decisions and the Psychology of Choice." In: *Science* 211.4481 (1981), pp. 453–458.

[264] J. Ulenaers. "The Impact of Artificial Intelligence on the Right to a Fair Trial: Towards a Robot Judge?" In: *Asian Journal of Law and Economics* 11.2 (2020), p. 20200008. DOI: doi:10.1515/ajle-2020-0008.

[265] J. J. Van Wijk. "The Value of Visualization." In: *VIS 05. IEEE Visualization*. IEEE. IEEE, 2005, pp. 79–86. DOI: 10.1109/VISUAL.2005.1532781.

[266] S. VanderPlas and H. Hofmann. "Signs of the Sine Illusion—Why We Need to Care." In: *Journal of Computational and Graphical Statistics* 24.4 (2015), pp. 1170–1190.

[267] E. de Visser, M. Cohen, A. Freedy, and R. Parasuraman. "A Design Methodology for Trust Cue Calibration." In: *Proceedings of the Intl. Conference on Virtual, Augmented and Mixed Reality (VAMR)*. Springer, 2014, pp. 251–262.

[268] C. W. Holsapple. "DSS Architecture and Types." In: *Handbook on Decision Support Systems 1*. Springer, 2008, pp. 163–189.

[269] W. E. Walker, P. Harremoës, J. Rotmans, J. P. Van Der Sluijs, M. B. Van Asselt, P. Janssen, and M. P. Krayer von Krauss. "Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support." In: *Integrated Assessment* 4.1 (2003), pp. 5–17. DOI: 10.1076/iaij.4.1.5.16466.

[270] E. Wall, L. M. Blaha, L. Franklin, and A. Endert. "Warning, Bias May Occur: A Proposed Approach to Detecting Cognitive Bias in Interactive Visual Analytics." In: *Proceedings of IEEE VAST*. IEEE, 2017, pp. 104–115. DOI: 10.1109/VAST.2017.8585669.

[271] E. Wall, J. Stasko, and A. Endert. "Toward a Design Space for Mitigating Cognitive Bias in Vis." In: *Proceedings of VIS*. IEEE, 2019, pp. 111–115.

[272] J. Wang, S. Hazarika, C. Li, and H.-W. Shen. "Visualization and Visual Analysis of Ensemble Data: A Survey." In: *IEEE TVCG* 25.9 (2018), pp. 2853–2872. DOI: 10.1109/TVCG.2018.2853721.

[273] M. Q. Wang Baldonado, A. Woodruff, and A. Kuchinsky. "Guidelines for Using Multiple Views in Information Visualization." In: *Proceedings of the Working Conference on Advanced Visual Interfaces*. ACM, 2000, pp. 110–119. DOI: 10.1145/345513.345271.

[274] J. Waser, R. Fuchs, H. Ribičič, B. Schindler, G. Blöschl, and E. Gröller. "World Lines." In: *IEEE TVCG* 16.6 (2010), pp. 1458–1467. DOI: 10.1109/TVCG.2010.223.

[275] J. Watróbski, J. Jankowski, P. Ziemba, A. Karczmarczyk, and M. Zioło. "Generalised Framework for Multi-Criteria Method Selection." In: *Omega* 86 (2019), pp. 107–124. DOI: 10.1016/j.omega.2018.07.004.

[276] J. J. van Wijk. "Evaluation: A Challenge for Visual Analytics." In: *Computer* 46.7 (2013), pp. 56–60.

[277] W. Willett, J. Heer, and M. Agrawala. "Scented Widgets: Improving Navigation Cues With Embedded Visualizations." In: *IEEE TVCG* 13.6 (2007), pp. 1129–1136. DOI: 10.1109/TVCG.2007.70589.

[278] G. J. Wills. "Selection: 524,288 Ways to Say "This Is Interesting"." In: *Proceedings of the IEEE Symposium on Information Visualization*. IEEE, 1996, pp. 54–60. DOI: 10.1109/INFVIS.1996.559216.

[279] D. V. Winterfeldt and G. W. Fischer. "Multi-Attribute Utility Theory: Models and Assessment Procedures." In: *Utility, Probability, and Human Decision Making: Selected Proceedings of an Interdisciplinary Research Conference*. D. Reidel Publishing, 1975, pp. 47–85. DOI: 10.1007/978-94-010-1834-0_3.

[280] K. Wongsuphasawat, Z. Qu, D. Moritz, R. Chang, F. Ouk, A. Anand, J. Mackinlay, B. Howe, and J. Heer. "Voyager 2: Augmenting Visual Analysis With Partial View Specifications." In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2017, pp. 2648–2659. ISBN: 9781450346559. DOI: 10.1145/3025453.3025768.

[281] J. Wood, P. Isenberg, T. Isenberg, J. Dykes, N. Boukhelifa, and A. Slingsby. "Sketchy Rendering for Information Visualization." In: *IEEE TVCG* 18.12 (2012), pp. 2749–2758. DOI: 10.1109/TVCG.2012.262.

[282] J. Wood, D. Badawood, J. Dykes, and A. Slingsby. "BallotMaps: Detecting Name Bias in Alphabetically Ordered Ballot Papers." In: *IEEE TVCG* 17.12 (2011), pp. 2384–2391. DOI: 10.1109/TVCG.2011.174.

[283] J. Wood and J. Dykes. "Spatially Ordered Treemaps." In: *IEEE TVCG* 14.6 (2008), pp. 1348–1355. DOI: 10.1109/TVCG.2008.165.

[284] X. Xie, F. Du, and Y. Wu. "A Visual Analytics Approach for Exploratory Causal Analysis: Exploration, Validation, and Applications." In: *IEEE TVCG* 27.2 (2021), pp. 1448–1458. DOI: 10.1109/TVCG.2020.3028957.

[285] C. Xiong, L. Padilla, K. Grayson, and S. Franconeri. "Examining the Components of Trust in Map-Based Visualizations." In: (2019). DOI: 10.2312/trvis.20191186. URL: https://diglib.eg.org.

[286] Z. Yan, R. Kantola, and P. Zhang. "A Research Model for Human-Computer Trust Interaction." In: *Proceedings 10th IEEE Int. Conf. On Trust, Security and Privacy in Computing and Communications, TrustCom 2011, 8th IEEE Int. Conf. On Embedded Software and Systems, ICESS 2011, 6th Int. Conf. On FCST 2011* (2011), pp. 274–281. DOI: 10.1109/TrustCom.2011.37.

[287] D. Yang, Z. Xie, E. A. Rundensteiner, and M. O. Ward. "Managing Discoveries in the Visual Analytics Process." In: *SIGKDD Explorations Newsletter* 9.2 (2007), pp. 22–29. DOI: 10.1145/1345448.1345453.

[288] M. J. Zaki and W. Meira Jr. *Data Mining and Analysis: Fundamental Concepts and Algorithms*. 2nd. Cambridge University Press, 2020. DOI: 10.1017/9781108564175.

[289] 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: *Proceedings of IEEE VAST*. IEEE, 2011, pp. 81–90. DOI: 10.1109/VAST.2011.6102445.

COLOPHON

This document was typeset using the typographical look-and-feel classicthesis developed by André Miede. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". classicthesis is available for both LaTeX and LyX:

https://bitbucket.org/amiede/classicthesis/

Happy users of classicthesis usually send a real postcard to the author, a collection of postcards received so far is featured here:

http://postcards.miede.de/

Final Version as of November 20, 2022 (classicthesis).