



Angeleitete visuelle Analytik für die Entscheidungsfindung bei Unsicherheit

DISSERTATION

zur Erlangung des akademischen Grades

Doktor der Technischen Wissenschaften

eingereicht von

Maath Musleh, MSc., BSc.

Matrikelnummer 12130534

an der Fakultät für Informatik
der Technischen Universität Wien

Betreuung: Associate Prof. Dr. Renata G. Raidou
Zweitbetreuung: Assistant Prof. Dr.techn. Davide Ceneda, TU Eindhoven

Diese Dissertation haben begutachtet:

Jean-Daniel Fekete

Anastasia Bezerianos

Wien, 27. September 2025

Maath Musleh



Guided Visual Analytics for Decision Making under Uncertainty

DISSERTATION

submitted in partial fulfillment of the requirements for the degree of

Doktor der Technischen Wissenschaften

by

Maath Musleh, MSc., BSc.

Registration Number 12130534

to the Faculty of Informatics

at the TU Wien

Advisor: Associate Prof. Dr. Renata G. Raidou

Second advisor: Assistant Prof. Dr.techn. Davide Ceneda, TU Eindhoven

The dissertation has been reviewed by:

Jean-Daniel Fekete

Anastasia Bezerianos

Vienna, September 27, 2025

Maath Musleh

Erklärung zur Verfassung der Arbeit

Maath Musleh, MSc., BSc.

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Ich erkläre weiters, dass ich mich generativer KI-Tools lediglich als Hilfsmittel bedient habe und in der vorliegenden Arbeit mein gestalterischer Einfluss überwiegt. Im Anhang „Übersicht verwendeter Hilfsmittel“ habe ich alle generativen KI-Tools gelistet, die verwendet wurden, und angegeben, wo und wie sie verwendet wurden. Für Textpassagen, die ohne substantielle Änderungen übernommen wurden, haben ich jeweils die von mir formulierten Eingaben (Prompts) und die verwendete IT- Anwendung mit ihrem Produktnamen und Versionsnummer/Datum angegeben.

Wien, 27. September 2025

Maath Musleh

Acknowledgements

First of all, I am deeply thankful for my advisor, Renata. She was a mentor and a friend throughout these four years. I cannot begin to describe my gratitude for her unconditional generosity and support. Without you, I wouldn't be here today. I owe you a debt of gratitude. Davide, I was lucky to have you as a co-supervisor on this journey. I learned and learned and learned from you. But most importantly, I learned how to develop the attitude of a researcher open to all feedback. And that is priceless. I thank Meister for his constant guidance and the productive research environment he fostered around us. I thank him for the serious intellectual discussions and the laughs. You are the heart and spirit of our research group! I thank Manu for all the scientific wisdom. You also made me a better teacher. I learnt a lot from you. I thank my doctoral colleagues for their support and knowledge. As I discreetly took notes, I learnt a lot from each of you, more than you know. Because of you, I am a better researcher. Daniel, Henry, Áron, Angeliki, Joey, Sarah...thank you! Hannah, thank you for sharing a desk with me and for all the laughs, talks, and support. Stephan and Simone, I do not think I've ever had such supportive colleagues. You never say, "later!", and that is something. Thank you! To Clarissa, Max, and Nicole, I know I've given you a hard time, but I sincerely appreciate everything you've done to make my time at TU Wien as smooth as possible. I also thank my colleagues with whom I always shared countless moments very early in the morning, without ever knowing their names; practicing my German, telling her about worries in my life and hearing about hers, the generous colleague who made sure we always worked in a clean environment; sharing a laugh with the security guard in the elevator; the other security guard with the fatherly smile greeting me every morning. Sorry for never knowing your names. But I hope you know how much you made a difference in my day. My teachers Matti and Geraldine, thank you for all the knowledge and wisdom. To all my colleagues and students whom I met throughout these four years at TU Wien, Aleksandr, Diana, Ildar, Lukas, Simon, Merut, Eric, and all colleagues I collaborated with in my papers, I have taken from you more than I've given. I thank you for your generosity with me. I thank all reviewers who thank every researcher who has ever reviewed my submitted work. I especially thank the reviewers of this dissertation, Professor Fekete and Professor Bezerianos, for their time and generous constructive feedback. If I unintentionally forgot any name, I hope you forgive me.

I thank my friends, who have all been my backbone and have all contributed one way or another to this dissertation. I thank Abir, Diana, and Aseel for the unconditional

support and encouragement. The 15 years of your friendship and camaraderie, through the hard times and good times, have a butterfly effect on every step I take forward. You will never begin to understand how much I learnt from you and how much you helped me grow, as a man and a Palestinian. Nada, you have unconditionally supported me and believed in me. Sorry, I was not there when I should have. And sorry if I will not be there if you lose more family in the genocide in Gaza. It is not this dissertation that kept me from sharing a tear with you. It is because we cannot make those who want us to capitulate feel that they beat us. But if we survive it, one day, when our people and our land are free, I will cry rivers with you for every family and friend we lost along the way, for every person we never got to meet because they were stolen from us. Special thanks to Brendan, who has always been a friend and a comrade. The discussion I shared with you, and every little thing I learnt from you as an academic and a researcher, as a human, helped me mature and grow. There are no words to describe how much I cherish our friendship. Thank you and Emma for the days we had in Ireland. It gave me the final push I needed to finish this dissertation. And I thank Bana, a person with unmatched intellect. But above all, a friend with unmatched empathy. All these debates that we had, all the disagreements, challenged me to think more deeply and clearly. To go for perfection, not an inch less. But I am sure you know how much your friendship impacted my life. And my gratitude to Haneen. Your friendship came at a crucial time as I started my PhD studies. You might not know how much you've given me, but it's definitely more than I can repay you. But I am also constantly ashamed that I could not be a friend you can lean on. My friendship fell deaf at the explosion of bombs! I am embarrassed that I got to complete my PhD, and this unjust world did not extend to you an equal opportunity. But particularly that, pushed me to continue and refuse the privilege of quitting midway. I do not know if you will survive to read this, as the killing machine is approaching your doorsteps, but I am sure that you will never be broken. And finally, my friend Bashar has given me perspective on life. You might not read this, but your friendship shaped me as a human.

I thank my parents for the unconditional love and support throughout these years. To my brother & sisters who have always been there for me, even when I distanced myself. I hope you know what you mean to me & how much I appreciate you. To my uncles, aunts, and cousins, who have been there for me, encouraging and supporting me. Thank you.

Fatima, you are my soulmate, the love of my life. Without you, I am not complete. You feed my hunger to knowledge. You make me feel safe. You make me a better man every day. You are the apple of my eye!

Last and foremost, I wouldn't have finished this dissertation if it weren't for those fighting daily for my liberty, defying all odds. Their brilliance on the frontline inspires me every single day. They knew how to employ their education and knowledge, fighting for the cause, the people, and this heaven, Palestine. One day, I will follow in your footsteps, as one is not complete without putting himself second to his people and land.

إلى الأرواح التي سُرِقَتْ منا، إلى الأرواح التي ضَحَّتْ من أجل
حُرِّيَّتنا في تلك العقد وعلى تلك الجهات، إلى تلك الأشجار
والشوارع والحجارة التي عَشِقْنَا ونَعَشَقُ كل يومٍ أكثر،
يؤلمنا الفقدُ والخسارة، ولكن يؤلمنا أكثر
أن نرفع رايةً بيضاءً قبل أن ننتزعَ مِنْ قَلْبِ الظلام
حُرِّيَّتنا التي لا نُساوم عليها، حُرِّيَّتنا التي لن ترضخَ للمستعمر.
لا قيمة لكل علمنا إن لم نسخره في قضية...ولا قضية تستحق
أكثر منك حبنا الأزلي فلسطين!

Kurzfassung

Visuelle Analytik (VA) ist aus der Notwendigkeit entstanden, die Entscheidungsfindung zu optimieren, indem menschliche Überlegungen in die Entscheidungsfindung einbezogen werden. Die Entwicklung von VA wurde durch bedeutende technologische Fortschritte in der Computergrafik und Datenverarbeitung erleichtert. Die Einbeziehung des Menschen in die Entscheidungsfindung zielt darauf ab, Hochrisikoszenarien zu bewältigen, bei denen automatisierte Ansätze der künstlichen Intelligenz (KI) unzureichend sind. Ein aktives Forschungsgebiet im Bereich der VA ist die Entwicklung von Methoden, die es dem Benutzer ermöglichen, effiziente und effektive Entscheidungen unter großer Unsicherheit zu treffen.

Das Zusammenspiel von Aspekten der Einstellung des Benutzers wie Vertrauen und Zuversicht unter Unsicherheit ist jedoch nicht ausreichend erforscht. Die Eigenschaften der Benutzereinstellung spielen eine entscheidende Rolle bei der Optimierung der VA-Entscheidungsfindung, sind aber schwer zu externalisieren und zu bewerten. So erweist sich beispielsweise das Vertrauen des Benutzers in seine Entscheidung als wichtiger Indikator für die Effektivität, wenn die Korrektheit der Entscheidung nicht gemessen werden kann. In dieser Dissertation untersuchen wir den Einsatz von Anleitungstechniken zur Bewältigung von Unsicherheiten bei der Entscheidungsfindung von VA, wobei wir uns auf Szenarien konzentrieren, in denen die Korrektheit von Entscheidungen nicht definitiv festgestellt werden kann.

In der ersten Forschungsarbeit, die in dieser Arbeit vorgestellt wird, schlagen wir eine zweidimensionale Anleitungsmethode vor, die die Detailorientierung und Eindringlichkeit der Anleitung berücksichtigt. Diese Methode wurde an einem VA-System zur Planung beispielhaft der Protonentherapie (PT) beispielhaft erprobt und mit Fachexperten bewertet. Die Evaluierung führte zu ermutigenden Ergebnissen, die das Potenzial von Anleitungen bei der Bewältigung hoher Unsicherheiten in der VA-Entscheidungsfindung belegen. Unsere Studie hat jedoch eine wichtige Herausforderung bei der Bewertung der Wirksamkeit von Anleitungen in solchen Szenarien aufgezeigt von Anleitungen. Das Vertrauen der Nutzer erwies sich als ein wichtiger Faktor für die Akzeptanz von Anleitungen durch die Experten. Wir haben außerdem eine Lücke in den Prüfverfahren identifiziert, die das Vertrauen der Nutzer als einen wichtigen Indikator für die Wirksamkeit von Anleitungen, erfassen können.

In einer zweiten Forschungsarbeit schlagen wir eine Methode zur effektiven Externalisierung und Messung des Benutzervertrauens in geführten VA-Szenarien vor. Die Methode geht darüber hinaus, sich auf das selbst angegebene Vertrauen zu verlassen, was erhebliche Einschränkungen aufweist, wie zum Beispiel kognitive Verzerrungen. Unsere Methode schlägt vor, die Herkunftsinformation in der VA als Netzwerkgraphen darzustellen, um Metriken aus dem Bereich der sozialen Netzwerkanalyse zur Messung des Benutzervertrauens einzusetzen. Dies wurde in einem Szenario aus der intelligenten Landwirtschaft mit einer Benutzerstudie getestet. Die Studie hat gezeigt, dass sich aus Netzwerkgraphen und sozialen Netzwerkanalysen objektive Maßstäbe für das Vertrauen der Nutzer ableiten lassen, die Anleitungen in der VA verbessern.

Schließlich stellen wir einen dritten Forschungsaufsatz vor, der sich mit der Herausforderung des Vertrauens der Nutzer in die Anleitungen selbst befasst, im Gegensatz zum zugrundeliegenden Vertrauen. Inspiriert von Arbeiten zu erklärbarer KI (XAI) schlagen wir ein Entwurfsmodell für erklärbare Anleitungen (XG) vor, welches das Vertrauen der Nutzer in VA-Lösungen mit Unterstützungen fördern kann. Das Modell wurde anhand von Anwendungsszenarien und eines Entwurfsdurchgangs mit einem Experten für VA-Anleitungen validiert. Das von uns vorgeschlagene Modell verbessert die Wirksamkeit von Anleitungen, indem es das Vertrauen des Benutzers in das VA-Werkzeug stärkt.

Während dieser Arbeit haben wir gelernt, dass ein mehrdimensionaler Anleitungsmechanismus effektiver mit Unsicherheiten umgehen kann, wenn diese schwierig zu quantifizieren und zu visualisieren sind, insbesondere im Fall von subjektiver Unsicherheit. Die Bewertung der Wirksamkeit von Anleitungsansätzen erfordert jedoch eine umfassendere Analyse des Zusammenspiels von Vertrauen und Zuversicht im Verstehensprozess. Die Verwendung von Herkunfts-Netzwerken und SNA-Metriken kann eine zuverlässigere und umfassendere Bewertung des Nutzervertrauens liefern, was darauf hindeutet, dass solche Ansätze zur Unterstützung ko-adaptiver Anleitungen eingesetzt werden können.

Abstract

Visual Analytics (VA) has emerged from the need to optimize decision making by involving human reasoning in sense making. The development of VA has been facilitated by significant technological advances in modern computer graphics and data processing capabilities. Involving humans in the loop aims to address high-risk scenarios where artificial intelligence (AI) automated approaches are insufficient. One active area of research with VA is the development of methods that enable the user to make efficient and effective decisions under high uncertainty.

Yet, the field of VA research has not fully understood how user attitude, namely trust and confidence, interplay in VA decision making under uncertainty. Properties of the user attitude play a crucial role in optimizing VA decision making, but they are challenging to externalize and evaluate. For instance, user confidence in their decision emerges as an important indicator of effectiveness when the correctness of the decision cannot be measured. In this dissertation, we explore the use of guidance techniques to address uncertainties in VA decision making, focusing on scenarios where the correctness of decisions cannot be definitively established.

In the first research paper presented in this work, we propose a two-dimensional guidance method that considers the guidance's detail orientation and intrusiveness. This method was exemplified on a proton therapy (PT) treatment planning VA system and evaluated with domain experts. The evaluation produced encouraging results demonstrating the potential of guidance in addressing high uncertainties in VA decision making. However, our study highlighted an important challenge in evaluating the effectiveness of guidance in such scenarios. User trust emerged as an important factor in the experts' acceptance of the guidance. We further identified a gap in evaluation approaches that can capture the level of user confidence, an important indicator of the guidance effectiveness.

In a second research paper, we propose a method to effectively externalize and measure user confidence in guided-VA scenarios. The method extends beyond relying on self-reported confidence, which presents significant limitations, such as cognitive bias. Our method proposes representing VA provenance as network graphs to exploit metrics from the social network analysis (SNA) field to measure user confidence, tested in a smart farming scenario with a user study. The study demonstrated that network graphs and SNA can derive objective user confidence measures that improve VA guidance.

Finally, we present a third research paper, which addresses the challenge of user trust in guidance, as opposed to confidence. Inspired by work on explainable AI (XAI), we propose a design model for explainable guidance (XG) that can foster user trust in guided-VA solutions. The model was validated with use case scenarios and a design walk-through with a VA guidance expert. Our proposed model improves the effectiveness of guidance by improving the user's trust in the VA tool.

Throughout this work, we learned that a multidimensional guidance mechanism can address uncertainties more effectively when uncertainties are challenging to quantify and visualize, especially in the case of subjective uncertainty. However, evaluating the effectiveness of guidance approaches requires a more comprehensive analysis of the interplay between trust and confidence within the sense-making process. Using provenance networks and SNA metrics can provide a more reliable and comprehensive assessment of user confidence, indicating that such approaches can be employed to support co-adaptive guidance.

التحليلات المرئية الموجهة في اتخاذ القرار في ظل عدم اليقين

ملخص

نشأت التحليلات المرئية (VA) من الحاجة إلى تحسين عملية صنع القرار من خلال إشراك التفكير البشري في عملية تفسير المعنى. وقد ساهمت التطورات التكنولوجية الهامة في مجال الرسومات الحاسوبية الحديثة وقدرات معالجة البيانات في تسهيل تطوير التحليلات المرئية. ويهدف إشراك البشر في هذه العملية إلى معالجة السيناريوهات عالية المخاطر التي لا تكفي فيها الأساليب الآلية القائمة على الذكاء الاصطناعي. أحد مجالات البحث النشطة في مجال التحليلات المرئية هو تطوير أساليب تمكن المستخدم من اتخاذ قرارات فعالة وكفؤة في ظل درجة عالية من عدم اليقين.

ومع ذلك، لم تفسر أبحاث التحليلات المرئية بشكل كامل كيفية تفاعل اتجاه المستخدم، وبشكل خاص الثقة والائتمان، في اتخاذ القرارات باستخدام التحليلات المرئية في ظل عدم اليقين. تلعب خصائص اتجاه المستخدم دورًا حاسمًا في تحسين عملية اتخاذ القرارات في التحليلات المرئية، ولكن من الصعب تظهيرها وتقييمها. على سبيل المثال، تظهر ثقة المستخدم في قراره كمؤشر مهم على الفعالية عندما لا يمكن قياس صحة القرار. في هذه الأطروحة، نستكشف استخدام تقنيات التوجيه لمعالجة حالات عدم اليقين في عملية اتخاذ القرارات في الذكاء الاصطناعي، مع التركيز على السيناريوهات التي لا يمكن فيها تحديد صحة القرارات بشكل قاطع.

في أول ورقة بحثية مقدمة في هذا العمل، نقترح طريقة توجيه ثنائية الأبعاد تأخذ في الاعتبار تفاصيل التوجيه ودرجة تدخله. تم توضيح هذه الطريقة على نظام لتخطيط العلاج بالبروتونات وتقييمها مع خبراء في المجال. أسفر التقييم عن نتائج مشجعة تثبت إمكانية التوجيه في معالجة درجات عدم اليقين العالية في اتخاذ قرارات في التحليلات المرئية. ومع ذلك، سلطت دراستنا الضوء على تحدٍ مهم في تقييم فعالية التوجيه في مثل هذه السيناريوهات. وقد برزت درجة ائتمان المستخدم كعامل مهم في قبول الخبراء للتوجيه. كما حددنا فجوة في مناهج التقييم التي يمكنها قياس مستوى ثقة المستخدم، وهو مؤشر مهم لفعالية التوجيه.

في ورقة بحثية ثانية، نقترح طريقة لتجسيد وقياس ثقة المستخدم في سيناريوهات التوجيه في التحليلات المرئية بشكل فعال. تتجاوز هذه الطريقة الاعتماد على الثقة المبلغ عنها ذاتيًا والتي تنطوي على قيود بارزة، مثل التحيز المعرفي. تقترح طريقتنا تمثيل بيانات مصدر التحليلات المرئية في شكل رسوم بيانية للشبكة لاستغلال المقاييس من مجال تحليل الشبكات الاجتماعية لقياس ثقة المستخدم، والتي تم اختبارها في سيناريو الزراعة الذكية من خلال دراسة المستخدمين. أظهرت الدراسة أن الرسوم البيانية للشبكة وتحليل الشبكات الاجتماعية يمكن أن تستمد مقاييس موضوعية لثقة المستخدم تعمل على تحسين التوجيهات في التحليلات المرئية.

أخيرًا، نقدم ورقة بحثية ثالثة تتناول التحدي المتمثل في اتجاه ائتمان المستخدم في التوجيه، في مقابل اتجاه الثقة. مستوحاة من أبحاث الذكاء الاصطناعي القابل للتفسير (XAI)، نقترح نموذج تصميم للتوجيه القابل للتفسير (XG) يمكن أن يعزز ائتمان المستخدم ناحية حلول التحليلات المرئية المدعومة بتقنيات التوجيه. تم التحقق من صحة النموذج من خلال سيناريوهات حالات الاستخدام وتجربة تصميم مع خبير في تقنيات التوجيه في التحليلات المرئية. يعمل النموذج المقترح على تحسين فعالية التوجيه من خلال تعزيز ائتمان المستخدم في أداة التحليل المرئي. خلال هذا العمل، استنتجنا أن آلية التوجيه متعددة الأبعاد يمكنها معالجة حالات عدم اليقين بشكل أكثر فعالية عندما يكون من الصعب قياس حالات عدم اليقين وتصورها، خاصة في حالة عدم اليقين الذاتي. ومع ذلك، فإن تقييم فعالية مناهج التوجيه يتطلب تحليلًا أكثر شمولًا للتفاعل بين الثقة والائتمان في عملية تكوين المعنى. يمكن أن يوفر استخدام شبكات المصدر ومقاييس الشبكات الاجتماعية تقييماً أكثر موثوقية وشمولية لثقة المستخدم، مما يشير إلى أنه يمكن استخدام هذه المناهج لدعم التوجيه التكيفي المشترك.

Contents

Kurzfassung	xi
Abstract	xiii
Contents	xvii
1 Introduction	1
1.1 Motivation	1
1.2 Vision and Research Questions	5
1.3 Contributions	7
1.4 Dissertation Overview	9
1.5 Authorship Statement	10
2 Overview and State-of-the-Art Positioning	13
2.1 Visual Analytics: The Genesis	13
2.2 Decision making in VA	15
2.3 Uncertainty in VA	21
2.4 Approaches for VA under Uncertainty	24
2.5 Guidance in VA	26
2.6 Guiding through Uncertainty	29
2.7 Evaluating Guidance	31
2.8 User Attitude in VA: Trust vs. Confidence	35
3 Uncertainty Guidance in Proton Therapy Planning Visualization	41
3.1 Introduction	41
3.2 Background	43
3.3 Task Analysis and Research Questions	46
3.4 Design and Implementation	47
3.5 Evaluation and Results	60
3.6 Discussion	68
4 ConAn: Measuring and Evaluating User Confidence in Visual Data Analysis Under Uncertainty	71
4.1 Introduction	72

4.2	Related Work	73
4.3	User Confidence vs. User Trust	76
4.4	Hypothesis and Research Questions	77
4.5	Dashboard for Farming as a Proof-of-Concept Scenario	78
4.6	Provenance Network Graph	84
4.7	Quantifying Confidence	87
4.8	Evaluation Methodology	90
4.9	Results	92
4.10	Discussion	96
5	TrustME: A Context-aware Explainability Model to Promote User Trust in Guidance	103
5.1	Introduction	104
5.2	Related Work	105
5.3	Explainable Guidance and User Trust	108
5.4	Research Questions	110
5.5	Methodology: A Model for Explainable Guidance	110
5.6	Refining the Model: Guidance Explainability	113
5.7	Ensure Applicability: Designing Explanations	127
5.8	Discussion	138
6	Conclusion	143
6.1	Reflection	143
6.2	Outlook	145
	Bibliography	149
	Curriculum Vitae	179

Introduction

1.1 Motivation

Computational methodologies are pivotal in augmenting human capabilities in data-driven decision-making processes, particularly as data volume expands exponentially alongside processing power advancements. Notably, Visual Analytics (VA) [WT04] approaches support human analysis by transforming the data into cognitively manageable information, enabling the user to interact with the visualization and underlying data to understand phenomena and make decisions in many fields, such as healthcare, sustainable farming, and business. The field of VA aims to foster data insight and promote knowledge discovery through an interactive loop between the user and a visual interface [Cui19]. The knowledge is generated through a sense-making loop that aims primarily to support users in answering high-level domain-specific questions [SSS⁺14].

For instance, VA can help stock market analysts decide on the best investment portfolio. They can assess the accuracy of their decision by pursuing investment in competing portfolios simultaneously and comparing future results to results obtained during the analytical process. In contrast, in radiotherapy (RT), doctors often analyze competing cancer therapy plans. However, they cannot validate their decision whether the pursued therapy plan is indeed the best, as it is not feasible to replicate the two competing therapy discourses on the same patient with the same conditions. The doctor can only evaluate the results of the chosen course of action. In both data comparison examples, a VA system may support the user by simulating and visualizing the expected results of alternative decision courses. VA techniques allow users to produce comparable insights that help them make a decision. In the first case, the possible alternative space is ample, i.e., the user can assess different combinations of stock options and amounts of investments in each stock. In the latter example, because of the computational complexity of simulating therapy plans, only a few alternative decision courses are compared—even if a larger combination of uncertainties can be explored to evaluate their effects on the possible

decisions. Moreover, in the stock investment scenario, users can validate the correctness of the decision, which feeds back into future decisions. However, in the clinical scenario, the inability to validate the absolute correctness of their decision limits its influence on future decision-making processes. These domain-specific features of the process impact the design of VA systems.

Often, researchers aim to optimize the decision-making process through the field of VA by proposing design approaches for interactive techniques that support the exploration and analysis of complex data structures [KKP⁺11, CG19]. These design decisions require a deeper understanding of the decision-making process, which is often driven by the user and contextual details within or across domains, such as user background, domain conventions, and cognitive biases, to name a few. These details can impact the understanding and conditions of an optimal decision—a *decision that produces the best outcome according to domain-specific value metrics*. To further clarify, we follow up on our two examples from the business and clinical domains. In business decision making, the risk of suboptimal decisions may not directly impact human lives but can have implications, such as financial losses.

Furthermore, business decision-making strategies, such as profit-maximizing strategies [SME08], shape the requirements for an effective and efficient VA system design. VA designers need to account for domain-specific challenges that impact the decision-making strategy, such as the *Winner's and Loser's Curse* specific to business decision making [SME08]. On the other hand, in the clinical domain, system designers combine computational and automated data analysis methods and interactive interfaces to support doctors exploring and understanding complex and heterogeneous data, such as electronic health records and medical images. VA techniques provide interactive graphical representations of anatomical, physiological, and pathological data to enable better-informed therapy decisions [RFG⁺20]. The accepted margin of risk in making therapy decisions is small because these decisions can result in significant harm to humans. Thus, several domain and contextual factors can impact the design and adoption of proposed VA systems, which makes the whole design, implementation, and use of VA systems quite challenging.

Besides decision making, VA can facilitate the discovery of new insights, the communication of findings, and the education of various audiences [CY13, PL20, KLL⁺22]. In other words, depending on the specific problem, it can sometimes be challenging to characterize the user's analytical tasks with precision [DS22]. This vagueness in defining user tasks complicates the design of effective solutions. These tasks can support the decision-making process of other persons who are not the direct users of the system. In such cases, system users are not the primary decision-makers, if at all. They become *proxy users*, mediating or moderating the decision-making process, in a scenario often referred to as shared decision making [BMMC23, BBK15, WCR20]; for example, doctors and patients decide on the optimal therapy plan together. This decision-making structure aggravates the challenges, introducing several biases and subjectivity into the process [BMMC23], which in turn further complicates the design. All these challenges, i.e., user and domain

constraints and task uncertainties, propagate and intertwine, resulting in a complex decision-making ecosystem.

On top of this complex ecosystem, decision making is often impacted by uncertainties. Scholars widely use the term *uncertainty* in biomedicine, meteorology, farming, and other domains to describe errors or ambiguities in the data and processes [TP21]. Uncertainty visualization is a field of study that aims to support the user’s decision making by communicating statistically modeled uncertainties [PKH21, HSB⁺22]. Techniques to visualize the uncertainties have been widely researched [TTvE16], aiming to improve decision making by raising the user’s awareness of the present uncertainties [MSH⁺23]. System designers deal with various objective and subjective uncertainties [GSWS21]. The former are irreducible uncertainties that refer to intrinsic statistical randomness, and the latter refers to various epistemic uncertainties that can be reduced or eliminated [CNdS07]. For example, in medical visualization, inherent uncertainties come from measurement approximations in data acquisition, noise during image scanning and reconstruction, bias in the image scanning and data collection methods, doubts in coordinate image registration due to organ or patient movement, and contesting processing methods, human interpretation, and calculation models [RPHL14, Rai18]. In stock market analysis, decisions are impacted by the uncertainty of future events [TF22]. Probable future values are assessed based on historical values that cannot always account for unexpected events, such as a pandemic. All these uncertainties can hinder the effectiveness of the decision-making process [DS22].

As seen in Figure 1.1, to inform decision making under uncertainty, VA designers propose techniques that can either 1) eliminate the uncertainties, 2) mitigate the uncertainties, 3) make the user understand the uncertainties, or 4) make the user aware of the uncertainties. These four strategies, respectively, provide adequate support to the user’s sense making, thus helping the user make an *optimal decision*. In the described medical field scenario, decisions to choose between alternative courses of action cannot be retrospectively measured for correctness, unlike, for instance, an application that supports decisions on choosing an optimal stock portfolio. In the latter example, one can measure the correctness of the decision by comparing it to the results of the alternative discourses; thus, the models of the VA system can be calibrated to support the decision making better.

When they cannot validate the correctness of their decisions, users increase their reliance on their background knowledge and experience, and existing biases in decision making [DFP⁺20]. Thus, more uncertainty is typically introduced to the process. As an example from the clinical domain, we take the field of RT, as it involves a complex decision-making process with significant implications for the patient. RT requires careful treatment planning to ensure that tumors are sufficiently irradiated without harming adjacent healthy tissues. In RT planning, a treatment plan computes how to optimally deliver the radiation dose to the patient. However, the patient data bears several layers of uncertainties [Rai18]. These uncertainties related to organ movement and radiation calculation models are unavoidable [FMCM⁺21], but they can be predicted and quanti-

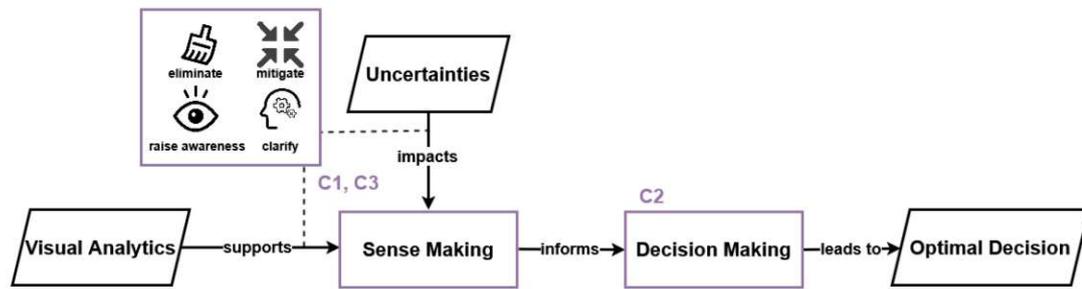


Figure 1.1: VA methods support users with the sense-making process that informs decision making intending to lead to an optimal decision. In the process, these VA methods can address the uncertainties that impact sense making by eliminating them, mitigating them, making users understand them (clarify), or making users aware of them. The chart simplifies the relationship of the sense-making processes to the decision-making process, which we discuss in more detail in section 2.2. As outlined in section 1.3 and located in [violet](#), this dissertation contributes a guidance mechanism (C1) and an explainability design model (C3) to support the sense-making process by addressing uncertainties in the analysis process. It also contributes a method to assess user confidence in VA decision making (C2).

fied. However, there is an uncertainty about which scientific model is best to adopt in calculating the uncertainties. Deciding on an optimal plan is a complex undertaking that must consider all the above-mentioned uncertainties. To account for the uncertainties, the doctor must rely on their prior experience and knowledge to decide on several parameters and models. This archetype of a high uncertainty scenario integrates rational and irrational decisions to optimize decisions that cannot be measured for absolute correctness. Such decision-making scenarios would profit from better-designed analytical tools that consider all aspects of the data and mend the knowledge gap of users [CGM19a]. An effective guided VA approach would converge the reasoning process by decreasing the perceived degree of uncertainty without compromising the trust and confidence of the user. In the RT scenario, trust relates to doctors' own belief that the planning system presents accurate and credible guidance. On the other hand, confidence relates to doctors' belief in the correctness of their sense making throughout the analytical process and the accuracy of their decisions. Both aspects are critical to the user's sense-making and decision-making process.

So far, we have discussed how VA supports decision making and how decision making under high uncertainty constitutes a particular challenge across fields. However, three challenges can impede the effectiveness and efficiency of uncertainty visualization. First, not all uncertainties can be visualized in all situations and contexts. For instance, uncertainty visualization techniques work best for uncertainties that can be quantified and statistically modeled [MSH⁺23]. Although objective uncertainties can be statistically modeled, not all subjective uncertainties can be statistically modeled. For example,

one cannot easily model the impact of a user’s attitude and cognitive biases on the analytical process. The second challenge is that visualizing uncertainties can impede decision making, for example, by negatively affecting users’ trust [SSK⁺16]. In such cases, visualizing uncertainties might not be an effective strategy. Other strategies are needed to boost the system’s trustworthiness, for instance, introducing guidance that presents the user with choices and recommendations to assist them in making the decision [Han21]. Third, visualization approaches can add more uncertainty to the analytical process and cognitively overload the user as uncertainties propagate and accumulate through all stages of the VA cycle [MSH⁺23], all the while, visual cluttering due to uncertainty encodings can result in cognitive overload.

VA can fall short in supporting the user’s reasoning under high uncertainty [Mac15]. Thus, we need a more robust and flexible solution that exploits a *lateral thinking paradigm*. Unlike vertical thinking, which is linear and systematic, lateral thinking exploits factors that drive *cognitive biases*, such as imagination, to create a more effective decision-making process under uncertainty [HV08]. With lateral thinking, users can *re-center their thinking* to address subjective uncertainties that originate from several sources, such as heuristics or cognitive biases [McK80, Poh17]. As their cognitive bias influences the decision-making process, users are often unaware of the uncertainties it produces or its impact on the process [DFP⁺20]. These biases, such as confidence biases, can yield decisional errors [Arn06]. Guidance refers to approaches in VA that allow the system to help users re-center their thinking during the visual analytics process, encouraging a lateral thinking mode. The literature proposes a need to investigate approaches to avoid the downside of visualizing uncertainties [Mac15].

1.2 Vision and Research Questions

In this dissertation, we explore *guidance* as a potential solution to address the challenges of VA for decision making under high uncertainty. Visual guidance can present an alternative to a comprehensive uncertainty visualization to support the user’s reasoning. *Optimizing decision making using guidance* has been successfully investigated in several fields, such as weather forecasting, material sciences, and economics [CGM19a]. However, as mentioned, applying guidance to optimize decision-making processes in some fields poses particular challenges due to domain-specific contexts, such as clinical conventions in the medical field and trust issues [PRSL23]. In many domain applications, the involved data and the decision-making pipeline include uncertainty, the impact of which needs to be conveyed to the user and considered when designing the guidance itself [RPHL14, GSWS21]. To this end, support for the comprehensive understanding of the decision-making pipeline and the accompanying uncertainty must be accounted for when designing guided-VA solutions. Guidance solutions can include strategies that help eliminate or reduce subjective uncertainties, as shown in Figure 1.1.

Subjective uncertainties are a source of knowledge gaps in VA processes. The *knowledge gap* [CGM⁺17b] in VA refers to the difference between what the user knows and what she

needs to know to complete an analytical task, hindering decision making. To efficiently and effectively complete analytical tasks, the user must comprehensively consider multiple layers of uncertainty and their impact on the analytical process.

As we discussed in section 1.1, domain-specific conventions and constraints complicate the uncertainty problem in the decision making throughout the VA pipeline. This problem impacts users' confidence in their decisions. Moreover, guidance tools designed for domain-specific contexts must first gain expert users' trust to support their adoption and use. Thus, probing available guidance design frameworks [HS23, SCEA23, CAA⁺20] must be context-specific to integrate effective elements that boost trust in the guidance tools.

Therefore, two critical values of the user attitude towards the proposed decision-making VA solution—*trust* and *confidence*—are anticipated to be impacted under uncertainty. A user-centered approach is crucial to guarantee the adoption of the solution, for example, by enabling users to verify the credibility and validity of the uncertainty computation models to trust them [PL20]. *Trust* and *confidence* currently exist in a confounding paradigm without clear conceptualization in literature [Smi05]—despite the two terms being critical concepts in designing practical guidance. Users' trust in novel methods, such as guidance, plays a vital role in ensuring the adoptability. At the same time, guidance tools must increase the user's confidence throughout the analytical process.

Still, VA solutions in some domains often *end up not being used* by their intended target audience. A potential explanation for this is that they are too complex to use as they integrate advanced visualization and analysis techniques that are not fully transparent to the users [PL20], i.e., the levels of user confidence and trust are insufficient. Supporting VA environments through guidance requires increased trust in the proposed approaches. Effective guidance assists the expert in navigating through uncertainties and making better sense of the data. Therefore, the analytical process may lead to a more confident decision, while we anticipate that confidence in decision making can increase decision precision.

Finally, the literature emphasizes the need to develop comprehensive evaluation tools to assess the effectiveness of visual approaches that guide users in decision making [Mac15]. In current work, researchers rely on self-reporting surveys [WAM⁺19] to evaluate user confidence, which is problematic as the users' self-confidence assessment is not always well-represented.

To respond to the challenges and problems discussed above, the proposed dissertation addresses the following research question:

RQ *To what extent can guidance-enhanced VA solutions support domain-specific decision-making processes under uncertainty?*

This general question branches into three sub-questions, which we aim to answer within this dissertation:

- RQ1** How can we design an *effective guidance* to boost experts' confidence in their decision making in the presence of uncertainty?
- RQ2** How can we effectively *evaluate the impact* of guidance techniques on user confidence?
- RQ3** How can we improve *users' trust* in guidance techniques?

1.3 Contributions

The research on guidance techniques for decision making remains sparse, particularly in high uncertainty scenarios, with no definitive ground truth and extensive comparison options [DS22]. Practitioners in many domains are often adamant about applying conventional decision-making processes. Thus, they have low trust in guidance and VA techniques that propose altering the conventional workflow. Typically, the complexity of the data, its inherent uncertainties, and the challenges of modeling it hinder the expert's trust in the reliability of the provided guidance. Furthermore, evaluation methods to assess the impact of the techniques on users' confidence rely heavily on self-reporting surveys. This limitation impacts the effective and efficient evaluation of the guidance techniques, especially their impact on user attitudes, such as confidence and trust.

Methodology To answer our research question (**RQ**), we begin by surveying the literature for guidance techniques and approaches. We base our understanding of guidance and related concepts, such as *knowledge gap*, on the work of Ceneda et al. [CGM⁺17b]. We recognize the specificity of decision making under uncertainty, where optimal decisions cannot be measured for correctness. Thus, we propose adapted definitions and conceptualization where needed.

To provide effective guidance, we must first understand the characteristics of the analytical scenarios in the field. Thus, as the first research endeavor, we collaborate with domain experts in RT to propose a guidance approach to improve therapy decision making under uncertainty in a clinical scenario (see chapter 3). We characterize the scenario in terms of data, users, and tasks [MA14], modeling the tasks based on well-known taxonomies in the field, such as Brehmer and Munzner's typology [BM13]. This characterization is needed to help identify and understand the uncertainties that must be addressed and is critical for successfully adapting guidance approaches and techniques. Furthermore, by collaborating with domain experts, we identified critical domain aspects, such as workflows and constraints, crucial for designing effective, context-specific guidance. This initial study allowed us to discover the emerging issues of user trust and confidence through the practical implementation of guidance solutions in a specific domain, as an *archetype of decision making in VA under uncertainty*.

The findings of the first paper inspired us to investigate further trust and confidence in guided VA. In a second research project (see chapter 4), we conducted a study to develop metrics to investigate user confidence throughout a complete analytical process. The

level of user confidence typically fluctuates and changes during the analytical process as it progresses. Thus, to represent and explore such fluctuations of confidence, we take inspiration from the field of graph and network analysis to develop a novel approach to represent and probe analytic provenance using network metrics. This research concurrently probes methods for improving user trust in guidance mechanisms, a critical factor for the successful development of guidance applications, particularly in sectors like medicine that are hesitant to adopt non-standard practices. We hypothesize that more trust in the guidance and its reliability improves user confidence, i.e., better guidance. Transferring results can be difficult due to domain-specific tool design [PCRHS18], so we evaluated our proposed metrics within an application domain that reflects a prototypical decision-making process, as discussed in section 1.1.

Finally, in a third research work, to complement the study of user confidence, we explore how explanations and explainability approaches can boost user trust in guidance (see chapter 5). We design and evaluate an explainable guidance (XG) model to design effective, trusted guidance in VA. We use an inductive thematic analysis to propose the characteristics of XG and the required components to consider when designing explainability. Our model maintains a space for the designer’s creativity in making design decisions. This approach helps balance the human and AI agency while outlining an effective XG design process.

Contributions This dissertation, along with the preceding research work, contributes to advancing the field of VA by proposing novel guidance-enriched VA approaches to support sense making and decision making under uncertainty as an answer to our posed research question (**RQ**). The work focuses on system→user guidance. This work presents and describes the following contributions, each of which provides a direct answer to the research sub-questions discussed above:

- C1** A framework for designing effective guidance approaches to facilitate the practical adoption of domain-specific applications.
- C2** Metrics that allow an objective assessment of user confidence.
- C3** A comprehensive model for explainable guidance that fosters user trust in guidance techniques integrated into a VA process.

We describe these contributions in chapters 3, 4, and 5.

Chapter 3, which reports on the study presented in our first paper [MMT⁺23], investigates uncertainty guidance mechanisms to support proton therapy (PT) planning visualization. We hypothesize that augmenting PT uncertainty visualization with guidance may influence the intended users’ perceived confidence and provide new insights (**C1**). To this end, the chapter presents an iterative co-design process with domain experts to develop a visualization dashboard enhanced with distinct level-of-detail uncertainty guidance

mechanisms. The proposed approach classifies uncertainty guidance into two dimensions: *degree of intrusiveness* and *detail-orientation*. The implemented dashboard supports comparing multiple treatment plans, i.e., nominal plans with their translational variations, while accounting for multiple uncertainty factors. The designed and developed strategies are evaluated by assessing perceived confidence and effectiveness during the sense-making and decision-making processes.

Chapter 4, which reports on the research presented in our second paper [MCER25], proposes a quantitative approach to measure user confidence — as opposed to trust — in an analytical scenario (**C2**). The paper illustrates how we exploit the respective user interaction provenance graph and examine the impact of guidance using a set of network metrics. We assessed the usefulness of the proposed metrics through a user study that correlates results obtained from self-reported confidence assessments and our metrics—both with and without guidance.

Chapter 5, which reports on the research presented in our third paper [MRC25], discusses the concept of *explainable guidance* and how it impacts the user–system relationship—specifically, a user’s trust in guidance within the VA process (**C3**). It subsequently proposes a model that supports the design of explainability strategies for guidance in VA. The model builds upon flourishing literature in explainable AI, available guidelines for developing effective guidance in VA systems, and accrued knowledge on user-system trust dynamics. The model responds to challenges concerning guidance adoption and context-effectiveness by fostering trust through appropriately designed explanations. We demonstrate the model’s value through two use cases. We also evaluated the model in a design walk-through with a guidance expert.

1.4 Dissertation Overview

This dissertation comprises research produced during the author’s doctoral studies between 2021 and 2025. We achieve the above-mentioned contributions in three papers published in top-tier journals in the visualization field.

The first paper, published in *Computers & Graphics* [MMT⁺23], is the base material for chapter 3, which proposes a guidance mechanism to mend the gap of adoptability into clinical practices, as an archetype domain (**O1**, **C1**).

Maath Musleh, Ludvig Paul Muren, Laura Toussaint, Anne Vestergaard, Eduard Gröller, Renata G. Raidou, “Uncertainty guidance in proton therapy planning visualization”, in *Computers & Graphics*, Volume 111, 2023, pp. 166–179, doi: 10.1016/j.cag.2023.02.002.

The second study, published in *Computer Graphics Forum* [MCER25], is the base material for chapter 4, which introduces novel metrics to measure *user confidence* (**O2**, **C2**).

Maath Musleh, Davide Ceneda, Henry Ehlers, Renata G. Raidou, “ConAn: Measuring and Evaluating User Confidence in Visual Data Analysis Under Uncertainty”, in *Computer Graphics Forum*, Volume 44, No.1, 2025, e15272, doi:10.1111/cgf.15272.

Finally, the third paper, published in TVCG [MRC25], is the base for chapter 5, which investigates a practical framework to design explainable guidance (**O3**, **C3**).

Maath Musleh, Renata G. Raidou, Davide Ceneda, “TrustME: A Context-Aware Explainability Model to Promote User Trust in Guidance”, in *IEEE Transactions on Visualization and Computer Graphics*, Volume 31, No.10, 2025, pp. 8040–8056, doi:10.1109/TVCG.2025.3562929.

Other papers published during the PhD, but not included in this dissertation:

Maath Musleh, Angelos Chatzimparmpas, Ilir Jusufi, “Visual analysis of blow molding machine multivariate time series data”, in *Journal of Visualization*, Volume 25, 2022, pp. 1329–1342, doi: 10.1007/s12650-022-00857-4.

The rest of the dissertation is structured as follows. Chapter 2 positions this dissertation in the state of the art. It presents a brief historical perspective on key concepts and ideas. It further untangles interchangeable concepts such as trust and confidence. Finally, after discussing the main contributions of this dissertation in chapters 3, 4, and 5. Chapter 6 concludes by reflecting on the work and presenting an outlook on future research.

1.5 Authorship Statement

This dissertation comprises research work authored by Maath Musleh (first author of the three papers below), who is the author of this dissertation for his doctoral studies. The work in chapter 3, chapter 4, and chapter 5 is the result of collaboration with various authors who contributed to this effort. In this section, we detail the contributions of our co-authors in each manuscript.

Paper I (Guidance for VA under Uncertainty) The first project was driven by a collaboration between my supervisor, Renata G. Raidou, and Aarhus University Hospital (AUH). It is the starting point that defines the archetype of the VA scenario that this dissertation addresses. This dissertation’s author and primary supervisor conceptualized the problem in collaboration with Ludvig Paul Muren, Laura Toussaint, and Anne Vestergaard from AUH. Professor Raidou and I designed the methodology to address the problem. Based on that, I designed and implemented the VA application with the support and supervision of Professor Raidou. All authors, except Eduard Gröller, validated the application through formative evaluation sessions. Dr. Muren, Dr. Toussaint, and Dr. Vestergaard curated the data used to test the application. Dr. Muren and Dr. Toussaint provided further domain-related resources, such as supporting data and studies, to assist the process. Professor Raidou and I

conducted the formal analysis of the study results and wrote the original draft of the paper. All the authors, except Dr. Vestergaard, conducted rigorous editing and reviewing. Professor Raidou acquired the funding and administered this project.

Paper II (ConAn: Analyzing User Confidence) This work is informed by gaps and limitations found in the first paper. The paper discusses methods to analyze user confidence in guided VA. My primary and secondary supervisors, Renata G. Raidou and Davide Ceneda, supervised the work. With their guidance and supervision, I was responsible for developing an approach to analyze user confidence during the analytic process and designing the user study. I conducted the user study sessions with twenty participants. Henry Ehlers validated and contributed to the design of the network-based metrics. I wrote the original draft of the paper, and all authors conducted rigorous editing and review.

Paper III (TrustME: Guidance Explainability Model) This project is the counterpart of paper II as we explore another close property of user attitude, user trust. I conducted an initial systematic literature review to collect concepts related to the explainable guidance (XG) model and design an initial draft of the model. These concepts and the models underwent a formative review and polish process by all three authors of the paper. I conducted the evaluation session with the guidance expert. Finally, I wrote the original draft of the paper, which was edited and reviewed by Professors Ceneda and Raidou.

Overview and State-of-the-Art Positioning

2.1 Visual Analytics: The Genesis

The roots of thought Figure 2.1 timelines the emergence of VA as a field of study. We can trace the inception of VA thinking in computer science and information systems to the field of cognition and perceptual studies [Got67] and *intelligence augmentation (IA)* [CN17]. These studies were thriving in the then-emerging digital image analysis in the early 1950s [Pre76], yet various fields adopted visual approaches in data analysis and decision-making. At the time, visual analysis emphasized analyzing the visuals, i.e., images [KJV⁺79]. Muller [Mul70] expressed enthusiasm at the potential of visual data analysis (VDA) empowered by technological advances. Thus, the 70s witnessed an era of active research to integrate graphical techniques into an *interactive* data analysis process [FFT74, Wel76]. Thereafter, research has witnessed a rise in investigations of methods that allow users to interact with the data during the analysis. This interest was further triggered by Tukey’s introduction of *exploratory data analysis* [Tuk77], while

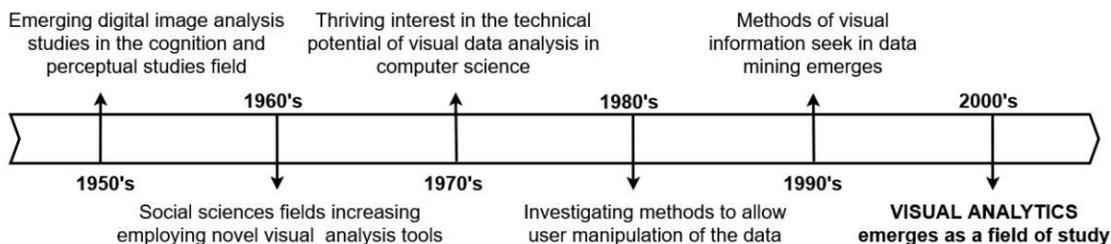


Figure 2.1: A timeline of the emergence of VA as a field of study.

Mizuta and Kawaguchi [MK83] proposed to make the data visualization *dynamic* under the user control. They presented *visualization* as a *system of analysis*, calling it “the motion of the points”.

However, VA formally emerged as an autonomous interdisciplinary field to study methods for analyzing data visually, i.e., using interactive visualization, with a stronger emphasis on the analytical aspect, in the early 21st century. The thinking behind it matured in the data mining field, where *visual information seeking (VIS)* curved a path to VA through Ahlberg and Shneiderman’s [AS95, p. 7] definition of VIS as “*presenting information visually and allowing dynamic user control through direct manipulation principles to traverse large information spaces and facilitate comprehension*”. The definition emphasized a user-centered interactive process. Soon after, Faconti and Massink proposed the idea of a “*continuous interaction*” [FM00] [Spe07, p. 144]. Thus, VA grew from integrating the user’s reasoning into the data processing stage [WP91, WC02] to integrating it within the complete analysis cycle. VA grew in importance as researchers realized that the user could engage more in visual analysis—not only through perception but also in actively reasoning and interacting with the visualization(s) [WGK10, P. 40]. VA could also provide solutions to the conundrum of the right amount of data to visualize (too little vs. too much data) [WGK10, p. 360].

Conceptualizing VA Eick [Eic00, p. 46] presented an interactive visual analysis framework, describing it as a “*discovery-based analysis [by] presenting information visually in an environment that encourages the exploration of linked events*” aiming for “*deeper insights and more actionable results*”. This description frames the process as a knowledge-discovery tool. It highlights two main connected goals: 1) insight discovery and 2) taking action. It also implicitly emphasizes the importance of linking multiple views as part of the process. Thomas et al. [TCK⁺01, p. 527] proposed a formal description of the visual analysis process as “*a two-way, interactive dialogue that enables the system to respond appropriately, learning from and aiding the human in the discovery process and allowing the user to develop and tell a story that represents the knowledge and learning that has been gained*”. They frame the process as a *dialogue* between the user and the system, highlighting the importance of an *adaptive* and a *mixed-initiative* system. In their description, Thomas et al. [TCK⁺01] focus on the high-level goal of communicating knowledge rather than making a decision.

Three years later, Wong and Thomas [WT04, p. 20] presented the first formal definition of VA as “*an approach to combine the art of human intuition and the science of mathematical deduction to directly perceive patterns and derive knowledge and insight from them.*” By defining it as an *approach*, they frame it as the study of the cognitive process in the visualization system. Visualization, as the host of VA approaches, is understood as the visual methods and tools that allow humans to obtain insight and detect interesting features and patterns in data efficiently and effectively by exploiting the unique capabilities of the human visual system [vW05].

Over the past two decades, VA’s definition has gone through several iterations. Keim et

al. [KAF⁺08, p. 157] elaborated on the definition of VA as a process combining “*automated analysis techniques with interactive visualizations for an effective understanding, reasoning, and decision making on the basis of very large and complex data sets*”. In a later study, Keim et al. [KMT10, p. 7] defined the process as the “*enabling and accessible analytic reasoning interactions supported by the combination of automated and visual analysis*.” Today, VA is understood as an approach of employing “*interactive visualization to integrate human judgment into algorithmic data-analysis processes*” [Cui19, p. 81555]. The definition emphasizes three main components: interactive visualization (the form/interface), human reasoning (the driver), and algorithms of data analysis (the engine).

To sum up, VA has emerged as an approach *to guide* humans in making more informed decisions [Han21, ZPF23]. Specifically, Cui [Cui19] emphasizes the VA process as an interactive sense-making loop, highlighting ongoing challenges in the collaboration and knowledge-sharing process. To overcome these challenges, researchers have proposed several solutions, for example, mixed-initiative guidance approaches from various streams that support the user in the VA process [CGM19a]. However, to design effective and efficient systems, it is essential to examine the decision-making process thoroughly. This understanding allows us to navigate the challenges of designing a VA system, which is ultimately framed as a visual decision support system, i.e., a system that guides decision-making. By reducing cognitive load, VA systems can help users arrive at more accurate conclusions [NSS⁺24]. Additionally, visual analysis enables users to externalize their thought processes [McK80], creating opportunities for adaptive reasoning and deeper engagement with the system.

2.2 Decision making in VA

The evolution of decision making The classic Hayy Ibn Yaqzan [Ibn09] illustrates how *the first human* gradually relied on their experience and observation to make decisions. Their biases and reasoning on primitive data and information gathered from their environment allowed them to improve their decision making. We visualized the processed data and information on the cave walls and tree barks, then on sheets formed from different materials. Visualizing this information was an artistic and lengthy process.

As *societies grew larger, human knowledge expanded, and statistical sciences evolved*, the decision-making process became even more complex. Scientists developed mathematical models to help them process uncertainties. Although impacted by potential biases, scientific modeling became the basis for analyzing data in many fields, making the decision-making process *more effective*. Decision making involves human reasoning, also known as sense making. *Sense making* is impacted by many factors such as experience, bias, and scientific models. With technological developments, scientists relied on several computational methods for data analysis to optimize the complex decision-making process by reducing error-prone sense making, offloading cognitive tasks, and mitigating human subjectivity and bias. This offloading has been achieved by automating many of the com-

ponents of the decision-making process, which reduces the length of human sense making. However, several ethical challenges arise with the automation of the process. Thus, researchers actively investigate methods to involve humans in the process [RBMK24].

Moreover, in *the industrial age*, the *data collection methods became more systematic*, yielding more diverse and more extensive data collected from all fields of life. New inventions, *in the computational age* in the early 20th century, allowed us to make the visualization process *more efficient*. The computational power supported data processing based on a scientific model [Chi20].

Humans integrated visuals and graphics to support their decision making long before the computer age [Fri08]. However, with the advancements in computational processing powers [Spe07, p. 6], interactive methods were incorporated to involve human reasoning more than before. New visualization methods became possible, such as patient scans in the medical field. Thus, visualization started playing a more significant role in the decision-making process. It also provided an alternative for computationally automated decisions to allow humans to understand and reason about decisions.

In *the information age* leading to the 21st century, the data collection abilities grew more significant with the progress of computational processing powers. This progress also allowed us to process and analyze more data. Visualizations were streamlined into the decision-making process by developing interactive methods to make the sense-making process more effective and to exploit the cognitive power of humans. VA as a field was born. As we highlighted in section 2.1, VA emerged as an extension of the visualization field, extending fields and concepts such as visual data analysis [Mul70], visual information seeking [AS95], and visual discovery and analysis [Eic00]. Methods in VA aim to integrate the reasoning power of humans with the computational powers to achieve a more effective and efficient decision-making process. Today, we witness tension between methods that aim to automate the decision-making process supported by the rapid development of AI methods and methods that aim to involve humans in the decision-making process using various visualization and VA approaches. This tension was evident in the EuroVis 2025 panel discussion “Visual Analytics at a Crossroads: Who Are We, and Where Are We Going?” [Lux25]. Novel approaches in our field are witnessing a push towards “non-stationary” tools, for example, in the study of virtual reality (VR) and immersive reality (IR). Challenges linger in defining the optimal balance between human involvement and computationally automating tasks in decision-making scenarios with high uncertainty. More recently, as we illustrate in section 2.5, visual guidance approaches have been proposed to support the user in the decision-making and VA tasks [DS22]. However, we still face fundamental challenges in designing efficient and effective VA tools. The human factors involved in the process, such as user trust and confidence, are not fully understood, exacerbating uncertainties in the sense-making process [PSBn23].

The sense-making dynamics Although researchers do not fully agree on how decision making relates to sense making [DS22], we anticipate that sense making in VA is part of

the decision-making process. The decision-making paradigm, seen in Figure 2.2, illustrates the two levels of the decision-making process: sense making and actual decisions. The sense-making process may demand that the person make several secondary decisions (*micro-decisions*), for example, choosing the parameters of a scientific model. As seen in the lower part of the pyramid in Figure 2.2, several factors drive the sense-making process, such as reasoning, scientific modeling, cognitive bias, emotional bias, and attitude [XAJK⁺15, NXB⁺16, Bol08]. This stage of the decision-making process can be supported by computational methods such as an interactive visualization or an artificial intelligence model. As seen in Figure 2.2, the computational method can support the user's sense making, and the user can also support it in return. For example, a smart farming VA system can help users make sense of crop data to make a decision on when to plant the crops. Users rely on their experience, which is a factor that influences their cognitive bias, which drives them to guide the VA system, for instance, by providing historic yield parameters to customize the visualization. Moreover, AI can support the interactive visualization method, and visualization methods can also support AI, for example, in explainable AI. AI can directly support decisions through automated approaches that cut out humans from the processes of making a decision to reduce cognitive load [NSS⁺24]. However, this approach presents several ethical challenges.

Finally, external factors, such as *uncertainties*, can disrupt the sense-making process and the decision. In the farming example, some uncertainties regarding the past rotational crops in certain beds can still affect their analysis. Users will reach a decision after they analyze the data in the system. They might still not be very confident in their decision due to some uncertainties, such as the impact of pesticides used in neighboring farms. They can also rely on integrated AI models to make decisions for them. VA draws on

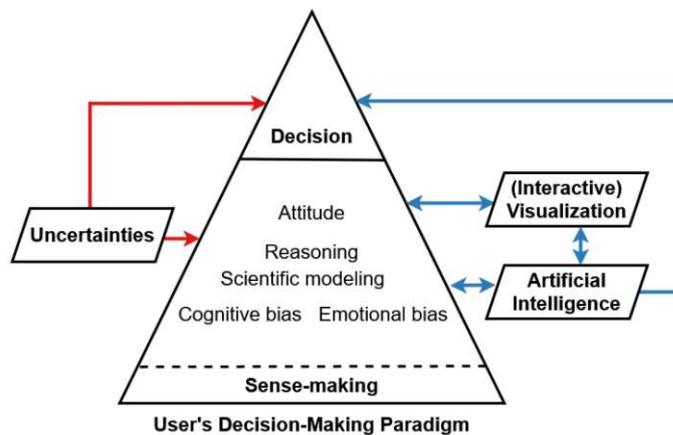


Figure 2.2: A decision-making paradigm. The decision-making process starts with sub-tasks that, within the sense-making process, are driven by several factors that lead to a decision. It is supported by computational methods such as (interactive) visualization and artificial intelligence. Uncertainties can also impact decision making.

expertise in areas such as human perception and cognition. Scholars have recognized the importance of the interactive human-machine collaboration to leverage the human cognitive potential in the analysis process, thus enhancing knowledge discovery [MK83, Pol83, TCK⁺01]. However, high uncertainty poses a challenge to effective and efficient decision making [MHTD24, CNdS07]. VA approaches may aid decisions under uncertainty by making the human decision-making process more effective and efficient.

The anatomy of decision making The literature describes two main decision-making modes: irrational and rational [NSS⁺24]. In Table 2.1, we list the most prominent conceptual synonyms used to describe these two types. An *irrational decision making* in VA refers to a process driven by tacit knowledge and rules-of-thumb, referred to as heuristics [NSS⁺24, CWK⁺17]. Also, user experience, emotions, and intuition contribute to irrational decisions. This decision making is appropriate for high-volume, low-impact simple tasks [WCR20], for example, routine decisions on daily restocking of medical inventory in a clinic. On the other hand, *rational decision making* refers to a process driven by explicit knowledge, complete awareness, and systematic reasoning [NSS⁺24]. It is more appropriate for high-impact, complex tasks where the available data is adequate [WCR20], for example, diagnosing a tumor of a patient. Svensson and Jacobsson [SJ24] extend it to a third type of decision-making mode, i.e., *action-based* mode. They describe it as “*doing in the situation*” [SJ24, p. 132].

Within the decision-making process, understanding, or sense making, can be bottom-up or top-down. In a bottom-up style, users analyze the details first to produce insights, known as verbatim knowledge [BMMC23]. This sense-making mode is also known as local analysis [GC02, BMMC23]. For example, smart farming VA users start by analyzing a heatmap of historic crops’ yield on their farm. On the other hand, in a top-down style, users start by making general ideas about the data, known as *gist knowledge* or the *bigger picture* [BMMC23]. It is also known as global analysis [GC02, BMMC23]. For example, farming VA users explore general trends of crop yield and weather patterns in the region.

However, the literature does not agree whether 1) choice (decision) *precedes* the action; 2) choice *produces* the action; or 3) choice aligns with action. In the latter case, our decision could be revealed through interpretation after we take action. Thus, making choices is seen as central to the decision-making process. Moreover, researchers disagree whether cognition produces emotions or precedes cognition [BMMC23]. A third view sees emotional and analytical thinking as complementary modes of thinking. Dhimi et al. [DT12] demonstrate how effective decision making is about the interplay of feelings (biases) and rational thinking, referred to as *quasi-rational* as described in the cognitive continuum theory. Although rational decisions might be more accurate, the user’s cognitive bias and heuristics, can be helpful in the decision-making process when tasks are constrained by time, during conditions that change rapidly, or where there is a limitation in the rational reasoning [BBK15, CWK⁺17]. Some research suggests that reliance on bias can improve decisions under uncertainty [GB09]. However, these biases introduce subjective uncertainties to the process that can impede the decision-making

Table 2.1: Different terms used to describe the two main types of decision making: irrational and rational. Irrational decision making refers to a process driven by experience and intuition, while rational decision making refers to a process driven by analytical reasoning and explicit knowledge.

Irrational	Rational
Emotional [SJ24]	Systematic [BMMC23]
Type I [Han21]	Type II [Han21]
System 1 [BMMC23]	System 2 [BMMC23]
Non-programmed [NSS ⁺ 24]	Programmed [NSS ⁺ 24]
Unstructured [NSS ⁺ 24]	Structured [NSS ⁺ 24]
Heuristic-based [WCR20]	Evidence-based [WCR20]
Experiential [SJ24]	Ostensive-based [SJ24]
Feeling [SJ24]	Thinking [SJ24]
Intuition [SJ24]	Cognitive [SJ24]
Affective [SJ24]	Higher-order reasoning [SJ24]
Holistic [BMMC23]	Analytic [BMMC23]

process if the user is unaware of them.

Moreover, in a realistic decision-making process, the human makes several *micro-decisions*, such as setting a model’s parameters, before completing the task and making the final decision [SJ24]. Also, the decision can be adjusted after interpreting an action and its consequences; thus, decisions are not firm [FM00]. They are continuous and can be changed during a single analytical process. In VA, systems may not be able to afford this circular process. Even when this is the case, researchers evaluate their VA designs based on an *analyze first, decide last* paradigm. As such, the results do not reflect many real case scenarios where the user might use the analysis to confirm a decision or an action, i.e., the analysis comes after the decision. The decision praxis is domain-dependent, where users can rely on different decision praxes. For example, Svensson and Jacobsson [SJ24] characterized eight different praxes in the medical domain.

The complex interplay of decision-making modes discussed above produces subjective uncertainties in the process. Heuristics, also seen as decisional shortcuts, can introduce several cognitive biases throughout the analysis process, such as anchoring bias and omission bias [BBK15, CWK⁺17, WCR20]. Decision making can be based on assessing a single course or comparing multiple competing courses of action. Definitions of decision making emphasize the comparative nature of the process [CWK⁺17, NSS⁺24]. However, not all decision options are discrete. For example, in a stock market VA tool, a user may want to decide how much money to invest in a particular portfolio. The decision output is of a continuous nature, i.e., monetary amount. Thus, in this dissertation, we distinguish comparative decision making as choosing between two or more discrete alternatives, for example, therapy plans A or B in clinical application. To simplify the complex uncertainty decision space, the user follows a multi-criteria decision model (MCDM) by breaking the

bigger task into smaller decisions, adding to the space of micro-decisions [NSS⁺24]. This decision-making scenario has been classified recently as a **CHOOSE** decision task by Brumar et al. [BMA⁺25].

Moreover, the comparative discourse decision making could be characterized by the capacity of future decision assessment, i.e., the ability to validate the accuracy of the decision. Thus, user knowledge uncertainty impacts the effectiveness of the decision [GMR⁺23]. VA approaches that assist comparative decision making aim to bridge the user’s knowledge gap. In such cases, the user uses the analytical task to choose between known options. With the presence of uncertainties, exploring the decision and micro-decision spaces increases our reliance on cognitive biases to avoid the clutter of perception—a subjective experience of an overcrowded knowledge space—due to our inherent tendency to simplicity [McK80]. Thus, we follow a *satisficing* strategy in our decision making, where we “*opt for a solution that meets minimum requirements under limited time and information constraints*” [NSS⁺24, p. 151]. This strategy can also impact the accuracy of our decisions.

Cognitive bias Cognitive bias is a crucial factor that can introduce uncertainties into the sense-making and decision-making processes. Cognitive bias is defined as “*a systematic deviation from norms or rational judgment, a way of creating a subjective-based reality based on an individual’s perceptions*” [Ker22, p. 101]. The literature identifies numerous types of cognitive bias, which Kerin [Ker22] groups into four major categories: hastiness, memory, information overload, and lack of knowledge. Dimara et al.’s [DFP⁺20] comprehensive taxonomy identifies 154 biases by task in information visualization. Dimara et al. [DFP⁺20] highlight how cognitive bias can lead to systematic errors in decision making. However, cognitive bias should not always be equated with error for two reasons. First, biases often lead humans to follow norms, not deviate from them [OWM08]. Second, the norm in judgments does not necessarily equate with correctness. Although confidence is present in this taxonomy, including more attitudinal-based biases, such as over-trusting, can improve our understanding of how cognitive biases impact decision making and introduce uncertainties.

To sum up, decision making is a continuously evolving process. To better understand decision making in VA, one must comprehend the nuances of how users make meaning of the data, visualization, and interactions. This sense making can be influenced by several factors that introduce uncertainties to the process—among which cognitive bias is an important factor. However, the field proposes a large number of biases that could affect the effectiveness and efficiency of decision making in VA. Thus, it is crucial to understand what uncertainties are introduced by these cognitive biases.

2.3 Uncertainty in VA

Types of uncertainty Uncertainty has been a field of intensive studies since the 19th century, when scientists developed mathematical models to reason with it. *Uncertainty* refers to events beyond definite knowledge that include several unknowns [PGA13]. It is not only a measurement of error—rather, a “doubt about the measurement result” [RPHL14, MSH⁺23]. Scientists classified uncertainty into objective and subjective [CNdS07]; Table 2.2 shows the different synonyms of these two types.

Objective uncertainty refers to the intrinsic randomness of events [CNdS07]. It is also referred to as stochastic uncertainty, a type of uncertainty that cannot be reduced. However, scientists addressed objective uncertainties by proposing statistical models that can make them better understood to facilitate their integration into the analysis process. Although scientific models were successful to a large degree in rationalizing objective uncertainty, subjective uncertainties remain not entirely understood [Gru24].

Subjective uncertainty is also referred to as epistemic uncertainty and involves uncertainties related to knowledge. These uncertainties can be eliminated or reduced when the knowledge gap is well-defined [CNdS07]. However, in most cases, subjective uncertainties are difficult to model. The gap in understanding subjective uncertainty is a challenge in the VA field, as designing effective domain tools can be impacted by high uncertainty that propagates and accumulates throughout the VA cycle. The uncertainty is not only present in the data but also in all other stages: the visualization, the modeling, and the user reasoning, as seen in Figure 2.3. To effectively address uncertainty in VA decision making, designers should investigate all perceived uncertainties present in the data, design modes to soundly process these uncertainties, and develop approaches that can efficiently and effectively integrate the modeling of the uncertainties in VA tools. Several challenges impede these three tasks.

First, uncertainties are user and domain-context dependent. Thus, it demands a closer investigation of the target user and their expected attitudes and behavior. It also requires a comprehensive understanding of the field and its peculiarities. Second, modeling uncertainties, especially subjective uncertainties, is challenging. In some cases, it is difficult to model uncertainty mathematically because of its intrinsic subjective nature. Third, several contexts, such as domain and user contexts, can impact the effectiveness

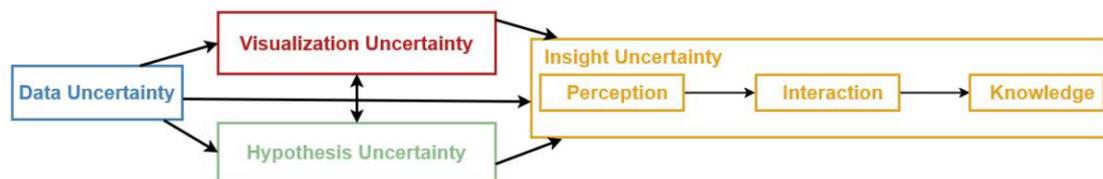


Figure 2.3: Uncertainties propagating through the VA cycle as proposed by the taxonomy of Gillmann et al. [GMR⁺23], highlighting the four components of the VA cycle: **dataset**, **hypothesis**, **visualization**, and **insight**.

Table 2.2: Various synonyms that describe objective and subjective uncertainties [CNdS07].

Objective	Subjective
Stochastic	Epistemic
Aleatory	Ignorance
Type I	Type II
Type A	Type B
Irreducible	Reducible

and efficiency of the uncertainty-aware VA design. Thus, we cannot design a one-size-fits-all tool. VA tools need to be responsive and adapt to diverse contexts. These contexts could also change within the process. For example, the user’s trust in the tool can improve or regress during the VA process.

In fields such as the clinical field, accurately communicating accumulating uncertainty is a challenge [Rai18]. However, the system must accommodate the perceived uncertainties to boost user confidence in sensitive decision making. A user perceives a fraction of all uncertainties present [Han21]. It depends on task requirements, domain analytical conventions, and uncertainty awareness limitations. These uncertainties could be universal or user-specific. The former refers to uncertainties intrinsic to the data and process, passed along at any stage of the processing and analysis pipeline. The knowledge gap could also be user-specific in that only some users might see or consider them, depending on the user’s experience and background knowledge. A gap may arise from different types of knowledge, such as domain knowledge or VA tool knowledge [CAA⁺20], amplified by the perceived uncertainty in the analytical space. These uncertainties can also aggravate the user’s bounded rationality. This concept refers to the human’s inability to make effective decisions due to limitations in processing all the available information [Bou08, MVM23]. Thus, guidance emerges as an approach that can assist the user by addressing some of these uncertainty challenges in VA decision making. It can alleviate some of the cognitive load and narrow down the large decision and micro-decision spaces. The user typically relies on subjective judgments to eliminate or select choices, which can introduce more bias and uncertainty to the process [NSS⁺24].

The compensatory strategy intrinsic to the MCDM, described in section 2.2, can be overwhelming as it “*aims to evaluate the alternatives by combining all information and consider the trade-offs between different factors*” [Han21, p. 16]. Guidance can support the MCDM strategy, making it more effective by reducing the reliance on subjective judgment and thus minimizing inconsistencies.

Sources of uncertainty Uncertainties are almost unavoidable in any field; such as predictions in weather forecasting [TTvE16]. The uncertainties propagated from imperfections of numerical weather prediction models to complexities in ensemble approaches are unavoidable [RBS⁺18, WHLS19]. Similarly, in agriculture, several uncertainties, such as

disease spread and weather uncertainties, impact the prediction of crop yields. These uncertainties emerge throughout the entire data collection and analysis pipeline. Moreover, they arise from several sources, such as missing data and its complexity [HKW23].

Bonneau et al. [BHJ⁺14] classify uncertainty sources into sampled data, modeling and simulation, and uncertainties in data and visualization processes. All these uncertainties, considered together, contribute to the complexity of the user’s *perceived uncertainty* [PGA13]. Gillmann et al. [GMR⁺23] propose a recent taxonomy of uncertainty events reflected in the four stages of the VA cycle. They categorize the events into six components: dataset, hypothesis, visualization, insight-perception, insight-interaction, and insight-knowledge. The latter three components are in the user insight stage of the VA cycle. The components reflect Pohl’s [Poh17] classification of cognitive biases into three types: memory, thinking, and judgment, respectively. Our work highlights a gap in classifying the uncertainty components proposed by Gillmann et al. [GMR⁺23]. First, it is particularly difficult to distinguish between the three user insight components. Pohl has also attested that the “*distinction between thinking and judgment [cognitive bias] may be less sharp...[and] memory illusions involve different kinds of thinking and judgment processes*” [Poh17, p. 3].

The task-based taxonomy of cognitive biases proposed by Dimara [DFP⁺20] can be used as the basis to expand the classification of uncertainties in visual analytics. Although a longer list of uncertainty components can overload the designer, it can provide an opportunity for more effective and efficient design decisions. Gillmann et al. [GMR⁺23] integrate some of the cognitive bias categories proposed by Pohl [Poh17] in other components on the system side; for example, the hypothesis uncertainty component in their taxonomy reflects the estimation task biases and hypothesis assessment biases in Pohl’s taxonomy [Poh17]. However, these uncertainties could also be extended to the user’s side. While the recall task bias is reflected in the insight-perception uncertainty component and the decision task bias is reflected in the insight-knowledge uncertainty component, Gillmann et al.’s uncertainty taxonomy [GMR⁺23] does not consider opinion reporting tasks. The five biases categorized under other tasks by Pohl [Poh17] are also worth investigating and integrating into the VA uncertainty taxonomy.

Finally, externalizing these uncertainties into the analysis process is an open challenge. Literature in uncertainty visualization proposes several techniques to effectively, efficiently, and soundly visualize uncertainty as an integral part of the VA. The uncertainties can be either displayed in a separate view or on top of other views, and they can be either static or dynamic [Mac15]. Unlike objective uncertainty, subjective uncertainty is difficult to quantify and model. Moreover, subjective uncertainties can interfere with other objective uncertainties. For example, consider a situation where a clinician uses several mathematical models to predict disease progression. Each model represents an objective uncertainty, based on statistical data and probabilities. However, the clinician’s decision on which model to use introduces subjective uncertainty, as it depends on their judgment, experience, and interpretation of the patient’s condition. Thus, with the presence of multiple competing models, displaying the objective uncertainty constantly can be

overwhelming and can clutter the user’s perception. Thus, a more flexible approach can facilitate the integration of uncertainty into the analytical process, namely, guidance. Guidance in VA can assist user reasoning under subjective uncertainty, a source of knowledge gap [CAA⁺20].

To sum up, uncertainties are critical to consider when designing visualization and VA tools. The two categories of uncertainties, subjective and objective, include numerous classified uncertainties. Several studies in VA contributed to classifications and taxonomies of VA-specific uncertainties. However, as the tools evolve, our understanding of uncertainties and how to include them in VA and decision-making processes keeps evolving. We must continuously explore methods to externalize and address these uncertainties efficiently and effectively.

2.4 Approaches for VA under Uncertainty

Addressing uncertainty has long been a central focus in the field of VA, with many works focusing on uncertainty detection, communication, modeling, and visualization [PWL97, JED⁺20]. *Uncertainty visualization* is the process of communicating uncertainty, such as accuracy and error, to support the user’s interpretation of data, models, or results [PKH21, JED⁺20]. This interpretation includes comparing potential alternative courses. Typically, techniques dealing with uncertainties are designed to bridge the user’s knowledge gap and reduce errors in decision making [KDJ⁺21], especially in a comparative context [BHJ⁺14]. For example, in weather forecasting, ensemble models account for vast alternative courses [TTvE16, WHLS19]. Hägele et al. [HSB⁺22] emphasize the need for probabilistic models to visualize uncertainties effectively.

Visualizing uncertainties Many researchers propose visualizing uncertainty to improve the accuracy of decisions in VA and the transparency of the analysis process [PKH21]. Padilla et al. [PKH21] map out the explicit and implicit visual approaches to communicate uncertainty through visualization. Focused on biological data, Weiskopf [Wei22] also presents strategies to visualize uncertainties effectively. These approaches and strategies are more suitable for addressing stochastic uncertainty using accepted statistical models in the domains. Furthermore, Padilla et al. [PKH21]’s study highlights several challenges in communicating uncertainty. Uncertainty visualization can mislead users by unintentionally suggesting non-existent categorical boundaries or deterministic information. Their study also highlights the user’s tendency to understand frequency framing better than probabilistic framing.

Padilla et al. [PKH21, p. 11] suggest visualizing uncertainty should obey the metaphorical norm that “*aligns with how users naturally think about data*”. However, this theory may assume a homogeneous user knowledge background. In reality, people decode semiotics differently based on their cultural knowledge base [Hal80, Hal97]. Uncertainty visualization designers cannot guarantee a *hegemonic reading*. Yet, *negotiated* or *opposed* readings can lead to misinterpretations in VA, and uncertainty visualization approaches

can be constrained by understudying the target user’s background or the context of developing a VA for users from diverse backgrounds.

Conventionality Researchers design visualizations to support various VA objectives. Battle and Heer [BH19] classify the systems based on the exploration goals and structures. In visualization, understanding the system’s exploration goals and structures converges. Battle and Heer state that a user could use a system with *a priori* formulated goals, or the goals could evolve during the process. They further set apart an open-ended, bottom-up exploration from a top-down process. Nonetheless, some systems could still incorporate a hybrid mode of exploration. For instance, a domain expert in fields such as clinical or farming arrives with set goals and a precise conventional exploration method shared across the field. Conventions in some domains present a challenge to designing and introducing new visualization approaches to the workflow. It should be done carefully while considering uncertainties propagating throughout the pipeline [Rai18, RFG⁺20]. A recent study by Peña-Araya et al. [PFW⁺18] highlights the malleable nature of uncertainty. Their work emphasizes the need to promote user agency when designing uncertainty visualization approaches.

AI for uncertainty Nazemi et al. [NSS⁺24] propose using AI approaches to support VA under uncertainty. AI approaches aim to address uncertainties by automating several stages of the decision-making process in VA. However, this approach can lead to transparency and trust issues. Human involvement in all decision-making stages is important in critical scenarios such as clinical decision making and scenarios characterized by traditional generational knowledge, such as farming [WJBC19]. Thus, in such scenarios in a VA system, user input in a mixed-initiative approach is crucial to optimize decision making.

Moreover, AI approaches may fail to address uncertainties propagating throughout the process. As Gillmann et al. [GMR⁺23] highlight, uncertainties can emerge at each stage of the VA cycle: dataset, modeling, visualization, and insight. The taxonomy of the uncertainty events in VA proposed by Gillmann et al. present a complicated picture that demands human intervention. Although this taxonomy is extensive, it can also be expanded to comprehensively capture more subjective uncertainties that impact the analytical process. A further elaboration on the insight uncertainty components is crucial to encompass different perspectives.

Quantifying uncertainty Maack et al. [MRP⁺25] propose a workflow to design uncertainty-aware VA. However, a crucial step in the workflow is detecting and quantifying the uncertainties. Most visualization approaches that communicate uncertainty are more suitable for addressing stochastic uncertainties. Stochastic uncertainty refers to quantifiable measures incomparable to ground truth, i.e., aleatoric uncertainty [GSWS21]. However, subjective uncertainties open up a vast knowledge gap that is difficult to integrate into the VA design, as it emerges from the rules of the process or disagreements in how a rule must be applied [RPHL14]. The challenge of quantifying subjective uncertainties

complicates the implementation of a guided VA solution. Moreover, we expand our understanding of subjective uncertainties to include qualitative uncertainties described by Panagiotidou and Moere [PM22]. Perceiving the subjective qualitative uncertainties propagated by the data is a challenge that requires a thorough investigation of the data transformation pipeline by the visualization researcher and domain expert [MSH⁺23].

To sum up, the absence of ground truth or models to comprehensively quantify some uncertainties limits the practical application of uncertainty visualizations [JED⁺20, HSB⁺22, BAOL12]. It is often challenging to capture all the uncertainties [MSH⁺23]. Thus, the system must prioritize the crucial uncertainties to address [Wei22]. These factors also limit the transparency and trustworthiness of automating several stages of the analysis process. However, we can present visualized uncertainties as guidance cues in a careful design to avoid a deceptive representation. These guidance are suggestions that a user can choose to follow or ignore.

Furthermore, in a mixed-initiative system, the user can provide input that improves the utility and effectiveness of guidance. Communicating uncertainties effectively is also crucial to improve the user’s confidence and build user trust [Rai18, SSK⁺16]. The mixed-initiative and adaptive nature of guidance can track user confidence and trust throughout the analytical process to adapt the guidance accordingly.

2.5 Guidance in VA

Guidance is often viewed from multiple confounding perspectives. Visualization and VA can be seen as guidance systems facilitating and assisting domain decision making [RBCR17]. Also, a user [GHG⁺19] or an ad-hoc VA [SSSEA20] may guide an AI algorithm to optimize a model’s results. However, guidance in VA aims to support sense making. The early work of Encarnaç o et al. [EaFB⁺94] highlights the need for system guidance at different visual analysis stages. Guidance can also support analytical reasoning, a topic that Dimara and Stasko [DS22] identify as a gap in developing VA. Guidance in VA can be thought of as a subsystem within the VA system because it deals with “*a distinct level of mental representation*” [BM99, p. 103].

The roots The Cambridge Dictionary defines guidance as “*help and advice about how to do something or about how to deal with problems*” [Cam24]. Since the mid-20th century, scientists have exploited the available computational power to guide decision making across various domains. This phenomenon fostered the growth of rich work in decision support systems (DSS) as interactive computer aids [KH79].

Guidance in information systems, such as VA systems, can be four-fold: user→user, user→system, system→user, and system→system guidance, as seen in Figure 2.4. This dissertation focuses on system→user guidance where the user is a human and receives guidance from the system, i.e., system guidance. *System guidance* is an expert system that can be constructed through *explicitly programmed* rules and operations or *inferences derived* from intelligent approaches such as machine learning [Gov23].

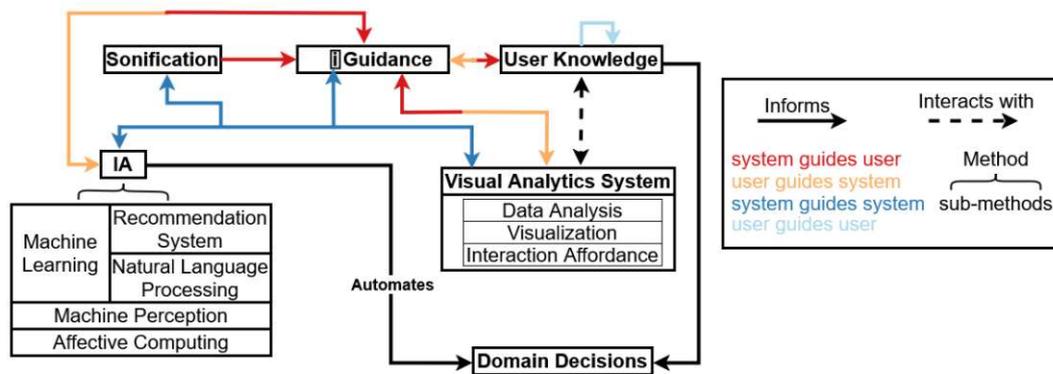


Figure 2.4: A general representation of the guidance role in the visual analytics process and its interaction with the main components. The guidance can be user-led or system-led. It interacts reciprocally with the user and the system. It can also be boosted with reciprocal AI or sonification approaches. Our work explores works that map to the red arrows in this illustration.

Initially, the system guidance encompassed five primary roles: storing and retrieving data, manipulating data, simulation, monitoring and feedback, and visualization [Har74]. With the computational advances in data analysis, visualization emerged as an essential guidance for human decision making [Mul70, Sil91, Dod08]. The field grew, benefiting from rich knowledge in several fields, such as human perception and cognition. As seen in Figure 2.4, a user can guide a user by feeding into the user’s knowledge. This accumulation of user knowledge from other humans and the VA system informs domain decisions. Domain decisions can also be automated.

Despite the advancement in the computer-assisted automation of the decision-making process, scholars noted the importance of the interactive collaboration between the human and the machine in the analysis process to enhance knowledge discovery [MK83, TCK⁺01]. As VA systems grew in complexity, they demanded more support for the users, and as a response, Ceneda et al. [CGM⁺17b] formalized guidance within VA. As seen in Figure 2.4, several methods such as AI and sonification can guide the user, via guidance, but they can also support the construction of intelligent guidance or the VA in a system→system guidance. The guidance tool, in itself, can also be guided by the user with direct parameterization. The user can also guide the AI model via guidance.

Conceptualizing guidance In 1991, Silver [Sil91] raised the importance of studying guidance in the information systems supporting human decisions. He defined *decisional guidance* as a process by which “a decision support system enlightens or sways its users as they structure and execute their decision-making processes” [Sil91, p. 107]. This definition frames guidance within the role of informing or influencing a user’s decision. It gave more weight to the system’s agency with less emphasis on the collaborative aspect of a guidance system. In the past decade, researchers invested substantial efforts into providing a

clear definition of guidance in VA [CGM⁺17b, CAS⁺18, SCW⁺22, HS23]. However, some confusion remains. Collins et al. [CAS⁺18] suggested that guidance can be used without a knowledge gap, i.e., to boost the user's confidence. As we explained earlier, the lack of confidence springs from uncertainty, and uncertainties are one source of the knowledge gap. Thus, a comprehensive and clear knowledge gap identification is essential to designing an effective guidance tool [CAA⁺20]. In highly uncertain decision-making scenarios, users require guidance to navigate complex models. Incorporating efficient and effective guidance techniques requires integrating them into conventional processes.

Ceneda et al. [CAA⁺20] propose a helpful framework to design guidance for clinical applications. They see guidance as a closed loop between the analyst and the system, informing each other. They outline six requirements for designing effective guidance. Nonetheless, the effectiveness requirement is often difficult to evaluate in VA scenarios where decisions cannot be measured for correctness. In such a context, the goal is to boost *user confidence* in their decision. This concept needs to be better defined and measured in the field. It is connected to the *user trust* and *trustworthiness* criteria, as discussed in section 2.8. Thus, these phenomena of user attitude should be investigated together when developing guidance to support VA.

The co-adaptive model proposed by Sperrle et al. [SCEA23, SJB⁺20, SJB⁺21] can boost users' trust in the guidance. The human-AI collaboration model proposed by Holter and El-Assady [HEA24] emphasize the adaptation component. Pérez-Messina et al. [PMCEA⁺22] extend the discussion of the co-adaptive model. They argued that system guidance should adapt to the user's complex behavior. User input in the co-adaptive model should be thought of as more than just the direct manipulation of guidance criteria afforded by the system. The understanding could be expanded to the indirect input afforded by the analytical provenance. Collecting and processing the provenance in real-time can improve the co-adaptive guidance model.

Positioning guidance within a larger Human-AI collaboration framework, Holter and El-Assady [HEA24] propose that the design space of guidance should consider the distribution of agency, the interaction intent and dynamics, and the adaptation methods. They view guidance as a co-adaptive mixed-initiative collaborative system. Their work highlights agency balance and users' cognitive processes as crucial factors in developing mixed-initiative systems, such as guidance.

To sum up, guidance remains a disputed concept in the study of VA. As the field grows, it is necessary to revisit its conceptualization, especially with the introduction of new AI approaches. Guidance in VA is viewed as a mixed-initiative co-adaptive tool. However, this dissertation focuses mainly on system→user guidance. We explore how it impacts and is impacted by properties of user attitude and how it can be employed to address uncertainties in decision making.

2.6 Guiding through Uncertainty

Guidance bridges the user’s knowledge gap by communicating uncertainties to the user. In communicating uncertainties, careful consideration of the field and target user is critical [SSK⁺16]. The effectiveness of the communication is improved by merging techniques and not relying on one technique only [PKH21]. Guidance techniques can support users in navigating, exploring, and interpreting uncertainty in data by providing cues, feedback, or recommendations. It is applicable only in the absence of VA methods to provide automatic answers using machine learning methods [CGM⁺17a]. Elmqvist and Klokose [EK25] emphasize that future research on guided-VA should focus on domain-customized guidance techniques that adapt to the users without taking away their agency. A recent survey of visualization tools by Oral et al. [OCW⁺24] suggests a gap in developing guidance tools to support users in VA.

Guidance and uncertainty Guidance may be used to address uncertainties throughout the VA cycle. These uncertainties can impact the user’s attitude throughout the analytical process. Enhancing uncertainty visualization with guidance can influence users’ confidence and ability to derive insights. Furthermore, firm adherence to conventions and predefined workflows, such as in the medical domain, impacts user trust and, subsequently, the adoption and use of system guidance. This mistrust towards the guidance may result from using very complex and unfamiliar visual representations or trying to alter domain experts’ existing workflow heavily, i.e., the guidance may recommend solutions unfamiliar to the experts. Thus, the problem emerges in VA on how to best integrate guidance into analytical processes while fostering user trust and confidence.

Guidance research still presents several challenges, with substantial gaps in transferable knowledge across the divergent streams of VA guidance research. Understanding how guidance functions and is used is crucial to converge the fundamental conceptualization of guidance in future work. Studies of guidance in VA are contradictory sometimes, but they maintain a rich diversity. An effective guidance tool can boost the system’s trustworthiness [Han21]. In exploring data with high uncertainty, a guidance-enriched system purposely disrupts the *status quo* of the expert’s analysis to add more confidence or minimize the low-level tasks leading to decision making [HS23]. The low-level tasks include the sense-making process, where knowledge is gained from the visualization. Sometimes, it can include micro-decisions taken throughout the analytical process before making a high-level decision.

For example, in the RT scenario in chapter 1, micro-decisions could relate to which calculation models should be used. In this scenario, exploration has a defined goal, i.e., deciding on the better plan from two options; however, there is no golden truth. The decision would rather depend on the expert’s reasoning. In such cases, two experts might arrive at two different plausible decisions. The expert expects the guidance tool to increase their confidence in their decision, choosing between two plausible courses.

However, evaluating guidance and its impact on user attitudes remains challenging. This

challenge presents a limitation for the effective design of guidance in VA. A further challenge that exacerbates the problem is the lack of metrics to assess the effectiveness of guidance. Our research investigates the problem of integrating guidance into VA to support decision making under high uncertainty. This problem poses significant challenges concerning achieving effectiveness without “harming” conventions while increasing confidence and trust.

Guidance strategies Designers can encode guidance in visual cues such as element stylization, glyphs, or contours. Furthermore, interactive guidance allows users to manipulate these cues by filtering plots and manipulating variables. Ceneda et al. [CGM19a] outline guidance techniques and approaches in the field. Most of the work reviewed provides orienting or directing system guidance [CGM19b]. Moreover, little work attempts to adapt the guidance based on the user’s indirect inference. The results point to the need for more dynamic guidance.

Early research in recommendation systems highlighted the value of adapting support to users’ intent. Gotz and Wen [GW09] proposed a method to infer user intent to improve visualization recommendations. The system predicts the user’s *visual analytic intent* based on detected patterns of analytical behavior. In recent work, Walch et al. [WSL⁺20] present an interesting example of exploiting provenance to provide effective “user-tailored guidance”. Their study shows promising results for provenance integration. However, the scalability of the approach poses a challenge. Group decision making (GDM) [HJCM22] is another interesting approach to designing guidance and uncertainty visualization techniques. Even if typically individuals make the decisions, aggregating an abstract of a user group can help design better techniques based on predictive models and collective knowledge. Ceneda et al. [CAGM22] employ a face detection technique to infer when the user needs guidance. However, they highlight the limitation of the agency of humans, who are not always guaranteed to trust the guidance.

Guidance can be generated using intelligent or explicitly programmed algorithms. AI methods are crucial in generating intelligent support for VA [BBR20]. Holter and El-Assady [HEA24] map guidance approaches within the space of a human-AI collaboration. Their work proposes the importance of investigating agency balance in the VA process. Although AI can improve guidance in many scenarios, explicitly programmed strategies should still be considered. A challenge lingers on how we define an intelligent agent. Monadjemi et al. [MGG⁺23] propose an agent-based framework to help analyze human-AI collaboration scenarios. Their work also highlights a gap in comprehensively understanding the user’s behavior. They propose that the human agent’s behavior should be represented as time series data because it changes constantly in VA scenarios.

Ha et al.’s [HMO24] study also focuses on AI guidance to VA. Their work investigates the challenges of AI-guided VA. They highlight transparency and trust issues that could arise in the process. However, their study does not map the impact of expertise on trust issues. Their work highlights that AI guidance may reinforce cognitive biases. This issue is crucial as it can propagate uncertainties in VA.

Moreover, Ha et al. [HMO24] explore the possible connection between transparency and trust in guided VA. This finding highlights the importance of considering the explainability of guidance. Ceneda et al. [CCEA⁺24] have also emphasized guidance’s explainability and transparency as an important quality in effective guidance.

VA literature presents many works on strategies to support the user in VA. Recent work [SCW⁺22, DWH⁺22] contrasts onboarding and guidance approaches. However, a broader perspective of guidance in VA may classify onboarding as a guidance approach. This topic remains an open debate. However, in VA work, we can identify several guidance approaches that are not explicitly presented as guidance but as aid or support approaches. Streamlining these approaches under the common guidance concept can enrich and support guided VA research.

To sum up, guidance is a subsystem in VA that can support the user in conducting the visual analysis by addressing uncertainties. This subsystem aims to boost user confidence in their decisions while maintaining trustworthiness and building user trust. Thus, measuring and evaluating the interplay of user attitude properties in the process is crucial to building effective guidance.

2.7 Evaluating Guidance

An adequate evaluation of the guidance tool is crucial to developing effective aiding tools in VA. However, Dimara and Stasko [DS22] identify a gap in methods to evaluate VA tools in the context of support decision making. The literature also highlights a gap in tools to evaluate approaches to address uncertainty in VA [KDJ⁺21, PKH21, WHLS19, HSB⁺22]. Researchers often lean towards using case studies to validate their designs’ and systems’ usability and usefulness [HLS⁺12, FNB⁺22]. The nature of uncertainty visualizations poses many challenges to a proper evaluation methodology [HQC⁺19]. Uncertainties are obstacles in scenarios where high-risk critical decisions are expected, while the intrusiveness of guidance may generate unwanted additional overhead [CAA⁺20]. Hullman et al. [HQC⁺19] categorize evaluation approaches in 86 papers that study uncertainty approaches. Their taxonomy identifies six levels that characterize these evaluation approaches. The survey highlights gaps in evaluating user attitudes in VA. Moreover, Hullman et al. highlight the tendency towards a confirmatory evaluation rather than an explanatory evaluation approach. This practice is particularly significant in evaluating approaches designed for scenarios where decisions cannot be measured for correctness.

Design studies It is a particular challenge to validate the impact of a proposed uncertainty visualization or guidance strategy on the decision-making process [QPCRM15]. The complexity of tasks under uncertainty in certain critical domains suggests that a design study is an appropriate evaluation method if it incorporates approaches, such as loud thinking, to understand what happens in the user’s head [SIBB11, SMM12]. The balance between knowledge acquired through available data and domain-expert knowledge is required to analyze visual results. Sedlmair et al. [SMM12] propose a nine-

step workflow for implementing design studies. The qualitative method's main objective is transferability. However, the concept of transferability is prone to interpretation. *Should the knowledge produced in the form of the visualization system be transferable to the domain, or should it be transferable to other domains?* The latter demands a higher level of abstraction than the former, with all the challenges that it entails.

Meyer and Dykes [MD19] build upon the ideas of Sedlmair et al. to propose an approach for a design study that treats it as a process of knowledge acquisition. In doing so, they borrow a perspective from the social sciences to emphasize the status of a design study as research work. They criticize the positivist approach to design studies and promote an interpretivist approach. As such, they stress the subjective nature of these studies and propose six criteria for validating a visualization design. They promote a rigorous design study grounded in a body of feedback knowledge. Their proposal stresses the importance of a reflective and complex process. Subsequently, they argue that the design study need not be reproducible, although they emphasize the importance of plausible evidence-based knowledge claims. The rigorous design study proposed by Meyer and Dykes [MD19] presents opportunities and challenges in validating guidance designs in uncertainty visualizations for critical scenarios, such as clinical and farming VA. A design study in some domains demands a well-rounded knowledge of conventional methods and visualization idioms familiar to practitioners. Resistance to change characterizes domains, such as the clinical field, where strict method conventions and the particular sensitivity of the confidence criteria (given the potential risks to patients' well-being) in the decision making pose a challenge to the transferability of knowledge. A design study that is not reproducible would threaten the trust and confidence criteria of the proposed design. Instead, a rigorous process involving the practitioners in the reflective tasks of the design study could help overcome the particular challenges of design studies in the domain.

Challenges and pitfalls A particular challenge we perceive is that the proposed method needs to indicate how we can distinguish the knowledge produced by validating the guidance techniques from that produced in validating the visualization system as a whole. The validation of both guidance techniques and uncertainty visualization systems is particularly challenging. Ceneda et al. [CGM⁺17b] address this challenge. They propose initial questions to direct thinking about decoupling guidance techniques from the visualization system in the evaluation process. In their paper, Ceneda et al. [CGM⁺17b] suggest classifying guidance techniques into three types based on the degree of assistance. They follow the classification with research that outlines the possibilities of employing the guidance in visualization. Identifying the knowledge gap is a driving factor in their decision tree [CGM⁺18]. Their work implies that this gap is only intrinsic to the user's knowledge. The method does not account for uncertainties arising from the intrinsic knowledge gap in the data. Nonetheless, the outlined decision tree provides a helpful tool for thinking about guidance in visualization.

Hullman et al. [HQC⁺19] present a taxonomy of evaluation methods of uncertainty visualization. In their survey, they noticed that "*subjective uncertainty is not a well*

defined construct in visualization evaluation” [HQC⁺19, p. 910]. They promote the inclusion of more subjectivity in participant responses. They also suggest that evaluation methods could be validated by comparing the decisions produced by evaluated systems with prior decisions. A challenge arises if the same user who has made a prior decision is the participant in the evaluation: *How could we account for the impact of prior decisions on post-decisions?*

Finally, Lam et al. [LBI⁺12] propose a helpful framework for customizing an evaluation. Their proposed approach helps a researcher better orient the evaluation methodology toward the questions of the research. In their work, they present seven evaluation scenarios that are linked to a clear line of questions. These were further enhanced by Isenberg et al. [IIC⁺13] with an additional eighth scenario. These scenarios can potentially be tailored to address different uncertainty cases, thus, providing a more flexible evaluation approach that takes into consideration the malleable nature of uncertainties [PFW⁺18].

Self-reporting metrics The available evaluation methods maneuver around the challenges of measuring the visualization impact on the final decision. Besides, the methods need to assess better the influence of the uncertainty factors on the process and sense making. Thus, we must evaluate the perception of accuracy and confidence achieved by guidance techniques. An effective guidance technique must boost the user’s confidence in sense making and decision making. Therefore, an effective measure of confidence is needed.

The intrinsic uncertainty in the evaluation methods exacerbates the uncertainty factors of the visualizations. These uncertainties could stem from a lack of a common understanding of certain concepts, study-context blind spots, a lack of understanding of individual knowledge gaps, or not knowing which aspects to measure and how to measure them. For example, *user confidence* is an important aspect to measure in studying guidance. Novel guidance techniques should be assessed for the impact on user confidence—primarily if the design affects decision making. We need metrics to quantify user confidence. However, *user confidence* is often confused with *user trust*. When clearly identified, researchers rely on self-reporting surveys to measure *user confidence* [WAM⁺19, NWHL20, PFCB23, PRJ⁺23]. This method emerges from the field of social psychology [SKJ15].

Ceneda et al. [CGM19b] design a user study to investigate the impact of the guidance degree on user performance. They also measure user feelings. Confidence is one criterion that they admit is difficult to quantify. Thus, as is conventional, they rely on self-reporting results. Furthermore, they note the limitation of conducting the study with non-expert users. It adds difficulties to validating the practical application of the guidance design. A transferable guidance study requires at least an entire analytical process evaluation session. In certain specialty domains, conducting a user study with experts is difficult, as the number of participants will be limited and will not produce statistically significant results. However, techniques can be abstracted and evaluated on more subtle application designs in fields with similar abstract decision-making processes. However, the scenario has to be accessible to a large pool of study participants.

Stasko [Sta14] proposes a value-driven evaluation that can be implemented in a user study to assess values, including confidence. Wall et al. [WAM⁺19, WXK22] extended Stasko's [Sta14] work to propose a heuristic-driven survey. Nonetheless, two issues linger. The first issue is the effectiveness of a self-reported response in assessing confidence. The second issue is using *confidence* to refer mainly to *trust*. Self-reporting surveys remain a limited tool to effectively investigate the guidance's impact on *user confidence*. It is difficult to guarantee a standard perception of confidence among all study participants, contributing to the evaluation method's uncertainty. This uncertainty is similar to the *comprehension uncertainty* described by Haghightakhah et al. [HEAF⁺22]. Quantifying the confidence value could support an effective evaluation, as it will provide an objective basis to understand it comprehensively.

Recently, Ceneda et al. [CCEA⁺24] proposed an eight-criterion heuristic that can be used to evaluate the effectiveness of guidance. The work outlines a clear protocol for conducting the evaluation. However, the evaluation also relies on the self-reported questionnaire. The suggested heuristics can be complemented with an approach to assess user attitude properties during the analytical process. This approach can present a more comprehensive picture of the impact of guidance on the analysis process beyond the direct user perception or recognition.

Analytic provenance Zhou et al. [ZAW⁺18] complement the self-reporting survey in their study with performance-based metrics to investigate the impact of uncertainty on *user confidence*. However, their approach relies on the *decision time* metric. More is needed to draw conclusions, as they point out in their study limitations. Boukhelifa et al. [BBC⁺20] noted the limitation of available metrics in evaluating uncertainty visualization. They address the issue by exploiting the insight carried by the analytic provenance. In their work, Boukhelifa et al. [BLB21] set apart *trust* from *confidence*. In their case study, they focus on measuring trust through analytic provenance. Their suggested metrics and approach could be extended to probe *user confidence*. Xu et al. [XOW⁺20] surveyed the use of analytic provenance in visualization research. Several studies exploit provenance to study visualizations. Eye-tracking [BKO⁺17] is one method used to probe the role of several visual elements in the analytical process. Smith et al. [SLMK18] also proposed evaluating user confidence in visual decision making by mapping eye gaze. Although this approach has considerable potential, eye-tracking [BKR⁺17, SLMK18] remains largely an expensive and inaccessible method. Few works used provenance in studying guidance techniques [BLB21]. Furthermore, the approaches surveyed by Xu et al. [XOW⁺20] did not propose metrics to probe and measure *user confidence* through analytic provenance.

Assessing user attitude Measuring user attitude remains an open question. Several previous approaches have been proposed over the years. Barki and Hartwick [BH94] proposed metrics to measure user attitude in information systems in work that pre-dates the inception of VA. However, as examples demonstrated previously, their metrics are extracted from self-reported questionnaires. On the other hand, Campos et al. [CNdS07]

proposed an approach to model subjective uncertainty [CNdS07]. Their approach can help capture uncertainties addressed by guided VA. These metrics can allow a better evaluation of the effectiveness of guidance approaches. User attitude could be considered a type or a source of cognitive bias. Thus, to investigate effective methods to measure user attitude, we can investigate prior work on cognitive biases. Arnott's [Arn06] taxonomy of 37 cognitive biases in decision making can be used as the basis to investigate how to develop a comprehensive metric system that captures the important biases that impact VA decision making under uncertainty. Cho et al. [CWK⁺17] use network graphs to externalize the anchoring effect, one of the cognitive biases. This approach of externalizing biases through interactive logs can be generalized to externalizing other cognitive biases. Xu et al. [XOW⁺20] present a helpful survey of methods that analyze provenance in VA. Two work categories are particularly useful: work that uses provenance to understand the user and work that aims to evaluate VA systems. Finally, Yanez et al. [YCON25] highlight the importance of adapting visualization to user needs. Their proposed taxonomy categorizes user input that can model user cognitive states to provide guidance as an intervention. Their work presents a helpful map to support design frameworks for guidance in VA.

To sum up, the VA field overrelies on self-reporting methods. These methods are insufficient and require complementary objective metrics that allow us to quantify user attitude properties. To develop effective guidance tools, it is crucial to develop accurate metrics to quantify user attitude as a source of cognitive bias. Analytic provenance can provide an input to methods to represent and progressively measure the changes in user attitude properties throughout the analytical process.

2.8 User Attitude in VA: Trust vs. Confidence

Developing and evaluating VA systems for decision making under uncertainty is challenging. To design an effective VA, the designer aims to understand the user's psychological state, i.e., attitude, that drives their analytical behavior. To validate VA approaches, the researcher needs to test them on realistic, complete scenarios to capture the realistic user attitude and behavior [BBK15]. However, while VA design itself influences user attitudes during typical analytical tasks, there remains a critical gap in understanding how different VA approaches affect users within specific domain contexts, e.g., in the medical field [BMMC23]. Designing a responsive VA that can adapt to the user's attitude during an analysis session is crucial.

User attitude in social psychology In VA, *User attitude* is the “*psychological state reflecting the affective or evaluative feelings*” towards a VA system [BH94, p. 62]. It refers to users’ “*affective evaluation of a system*” [BH94, p. 59], or in other words, the user's judgment of the quality of a system and its components. Such an attitude, which is a context-dependent state, may fluctuate several times during an analytical session and may also be influenced by several cognitive biases. User attitudes can also impact user

behavior and the effectiveness and efficiency of VA solutions [BBK15]. For instance, this impact can happen through what is known as *bounded rationality*, where user context plays a big role in their decision making. For instance, in economic studies, users can be split between *satisficers* and *maximizers* [MVM23]. If the user is a researcher, they might lean towards being maximizers who are more interested in the analytical process. On the other hand, practitioners lean towards being satisficers who are interested in the result being “good enough”.

User attitude is more challenging to assess than user behavior because it is more difficult to observe. User behavior tends to manifest in explicitly observable user actions. On the other hand, user attitude takes place within the user’s cognitive psychology, making it challenging to examine. However, in VA, attitude can manifest as a *pattern* of analytical behavior that reflects a *mental organization* [You32] [Bal01, pp. 88–89]. This organization—if known or measurable—allows VA designers to improve design decisions.

User attitude consists of multiple properties that can impact each other. Confidence and trust are two major properties that have been targets for recent VA research. These two terms are often used interchangeably in VA. However, the logical semantics of the two concepts are different. A better understanding of their logical and lexical semantics helps the development of effective evaluation metrics to measure confidence and trust in VA systems. The two concepts are a *social psychology construct*.

Returning to the RT example discussed in chapter 1, we aim to provide a departure point to understanding and distinguishing between both concepts. A medical physicist uses a treatment planning system (TPS) to determine the most appropriate plan for her patient. She initiates the task with a knowledge base that defines her *trust* in the system, the correctness of algorithms used to calculate a certain conversion factor, and the accuracy of a histogram plot and the dose overlay slice visualization. She also *trusts* the underlying scientific methods used for these calculations. Preceding the analysis process, she also has a certain degree of *confidence* in finding the most appropriate solution because she *trusts* her knowledge and expertise. Her *trust* in the system, data, and information presented also impacts the degree of *confidence* she possesses. As she proceeds with the data exploration, she makes exploratory decisions/choices with a certain degree of *confidence*. Eventually, she chooses the optimal plan based on her assessment. She is somehow *confident* about her decision, but with high doubts as the uncertainties are daunting.

As shown with the example above, the logical semantics of the two concepts—*trust* and *confidence*—are different. Sociology distinguishes both concepts by the relationship to *agency*, the presence of *risk*, and *consequence* [MW09]. Contrary to trust, according to Luhman [Gam88, Luh05], confidence gives out agency, “a feeling of control over actions and their consequences” [Moo16]. In the example above, the physicist trusted the integrity of the visualized data. She does not have strong evidence of the completeness of the data. Nevertheless, she chose to trust the system. Thus, she retains her agency. Without evidence and with high uncertainties, she tolerates a high risk. However, when she made her final decision, it was based on the given evidence by the system. The risk is lower

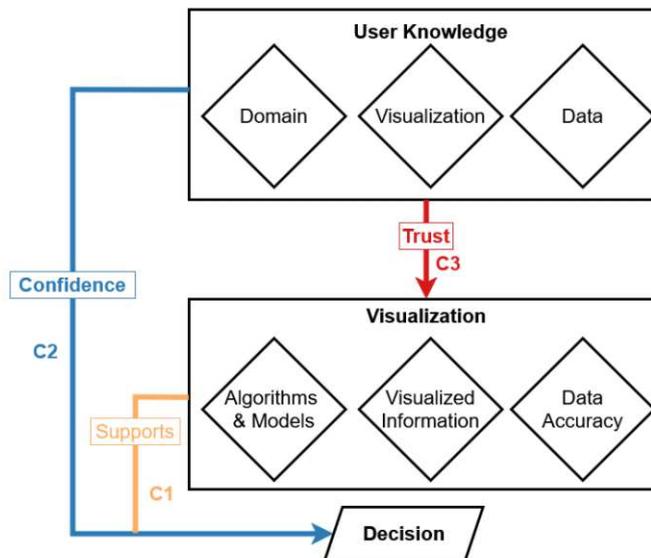


Figure 2.5: The semantics of *trust* and *confidence* within an analytical process. Confidence and trust are based on the user’s starting knowledge. The visualization supports users’ confidence in their decisions with all its components. C1, C2, and C3 reference the area of our contributions as outlined in section 1.3.

than in the previous case. Thus, the agency, in this case, is given over to the system. The error in the decision is, then, attributed externally, i.e., an error in data or missing data. On the other hand, in trust, the blame is internal, i.e., one’s analytical reasoning. A better understanding of the logical and lexical semantics of trust and confidence helps the development of effective evaluation metrics to measure confidence and trust and, subsequently, design more effective guidance techniques. In Figure 2.5, we illustrate the semantics of trust and confidence within an analytical process.

Confidence is present in the decision-making stage of the analytical process, where the acquired knowledge points out one option. The error in the decision is, then, attributed externally, i.e., error in data, missing data, etc. On the other hand, in trust, the blame is internal, i.e., one’s analytical reasoning. Trust requires a risk situation where uncertainties are high. The user chooses to trust that the information provided by the system (in terms of visual encoding in VA systems) is reliable. It could refer to trust in internal algorithms and calculations embedded under the hood.

In psychology, confidence is seen as the self-perception of one’s decisions. It is usually assessed by a self-reporting method. Confidence also refers to the “*judgments of accuracy...after the completion of a task*” [SKJ15, p. 158]. In the field, there are several tools to measure performance-based confidence. For example, the Proverbs Matching Test [SKJ15] could be adapted to measure self-perceived confidence in decisions within the analytical process. Lack of confidence is not the only challenge to effective decision

making. Overconfidence is a challenge that arises from confidence bias [Arn06].

On the other hand, trust is understood as one's perception of someone or something else. Trust in psychology is assessed using similar self-reporting methods. It can be more accurate than assessing confidence. However, the challenge arises in examining the reasons for the trust or mistrust. Researchers in psychology disagree whether trust is a rigid state learned early and remains stable throughout the relationship cycle or malleable and continuously changing [Sim07]. This dissertation approaches trust as a malleable state [PG15]. However, it assumes that the early stage of the user-system relationship is crucial to establishing a trusting relationship. Trust is also understood as a *psychological contract*, “an individual's beliefs about the terms and conditions of a reciprocal exchange agreement” [Rob96, p. 575]. Thus, trust deteriorates when expectations are not met. Although over-trusting can affect decision making, it can be overcome by evaluating how effectively the other party met the expectations, or *terms of contract*.

Trustworthiness emerges as an important related concept in understanding and evaluating the trust relationship, especially in ethics studies. Trust is to trustworthiness what reliance is to reliability. Trustworthiness is viewed as the moral obligation of the trusted party to act with goodwill in fulfilling expectations under the psychological contract [Jon12]. For example, in VA, the user expects the system to provide an accurate and comprehensive encoding of the data. The importance of trustworthiness in the relationship between two parties, e.g., the user and the system, is that it holds the trusted party accountable and responsive in the relationship that is essentially based on high uncertainty and risk, as we stated above [Jon12].

Confidence vs Trust in VA Confidence and trust are properties of an attitude towards the self and the VA process. In VA, this attitude is based on knowledge that exists and is acquired in the process. Users actively seek information that informs their attitude toward the self (confidence) and the VA tool (trust). This information represents the knowledge gap that needs to be bridged by VA tools, i.e., guidance.

User confidence could be understood from two different perspectives. It could refer to self-efficacy or the belief in one's accuracy [SKJ15]. *Self-efficacy* is the starting point of confidence that guidance must boost throughout the analytical process by supporting users' belief in the correctness of data and information interpretation. Moreover, it should increase the user's belief in the accuracy of their choices throughout the analytical process. On the other hand, *user trust* is the belief in the correctness of an external party without evidence [Cof07]. In VA, trust exists within the interaction between the human and the VA system. Guidance designers should consider how their design choice impacts the user's trust in the tool and their confidence during the analytical process.

In our work, *confidence* refers to users' degree of certainty in the correctness of their decision based on specific acquired knowledge. While *trust* is the users' degree of certainty in the system's information/decisions/options. It is a holistic impression of certainty in the space of uncertainty. When users trust, their actions are interdependent on the system rather than the self-accuracy of judgment, such as confidence.

Thus, measuring user trust in a system or data could be more complicated, as it requires probing several sources of knowledge at a broader scale. This assessment varies between users based on their backgrounds and expertise. Trust is less likely to be impacted during one visual analysis process. As discussed above, the early stage of the relationship is crucial. However, confidence is a measurable concept and highly mutable within an analysis process. We highlighted in section 2.2 how properties of user attitude can be a source of cognitive bias, which can result in subjective uncertainties. The user context and expertise influence such biases. For example, the Dunning-Kruger effect describes how “*low-ability people overestimate their performance*” [DFP⁺20, p. 18]. This effect is closely related to the properties of users’ attitudes. The Dunning-Kruger effect, a cognitive bias, can result from overconfidence.

To sum up, in our dissertation, we distinguish two important properties of user attitude: user confidence and user trust. This distinction allows visualization and VA designers to tailor their proposed techniques better to address uncertainties more effectively. Understanding the interplay between both properties is also important. This comprehensive understanding allows systems to adapt progressively to the changes in the properties of the user attitude throughout the analytical process. It also allows the development of more accurate tools to quantify these properties, allowing a more effective evaluation of the designed tools. Finally, the distinction between the two properties generalizes across visualization and VA studies. This chapter exemplifies guidance as the object of trust and a VA tool to support confidence.

Uncertainty Guidance in Proton Therapy Planning Visualization

This chapter addresses the need for guidance mechanisms to support decision making in a clinical domain, namely, proton therapy (PT) planning, and is based on the publication:

Maath Musleh, Ludvig Paul Muren, Laura Toussaint, Anne Vestergaard, Eduard Gröller, Renata G. Raidou, “Uncertainty guidance in proton therapy planning visualization”, in *Computers & Graphics*, Volume 111, 2023, pp. 166-179, doi:10.1016/j.cag.2023.02.002.

Uncertainties in the PT workflow pose significant challenges for navigating treatment plan data and selecting the most optimal plan among alternatives. Although guidance techniques have not yet been applied to PT planning scenarios, they have successfully supported sense-making and decision-making processes in other contexts. We design a multidimensional guidance mechanism to support PT decision making under uncertainty. Our findings indicate that uncertainty guidance in PT planning visualization does not necessarily impact the perceived confidence of the users in the process. Nonetheless, our work provides new insights and raises uncertainty awareness during treatment plan selection. This observation was particularly evident for users with longer experience in PT planning.

3.1 Introduction

Proton therapy (PT) is a standard radiation modality in cancer treatment. It requires careful planning to ensure that a tumor will be sufficiently irradiated while adjacent tissues are avoided as much as possible. The treatment plan is calculated using a dedicated treatment planning software (TPS), which computes how the therapy system will deliver

3. UNCERTAINTY GUIDANCE IN PROTON THERAPY PLANNING VISUALIZATION

the radiation dose to the patient. This lengthy process is usually limited to generating a couple of alternatives with additional positional, i.e., translational, variations. Deciding on an *optimal plan* is a complex undertaking with several uncertainty factors, which relate to the physics behind the calculations and the biological effects of the dose on tissues.

Researchers and practitioners design and select robust PT plans for patients, depending on the available TPS to calculate the plan(s) and make appropriate therapy decisions. The TPS includes mainly slice-dose overlay views (Figure 3.1 (a)) and an additional plot, called *dose volume histogram* (DVH), that depicts radiation dose administered to volume percentages of specific structures (Figure 3.1 (b)). Beyond juxtaposition, the TPS views

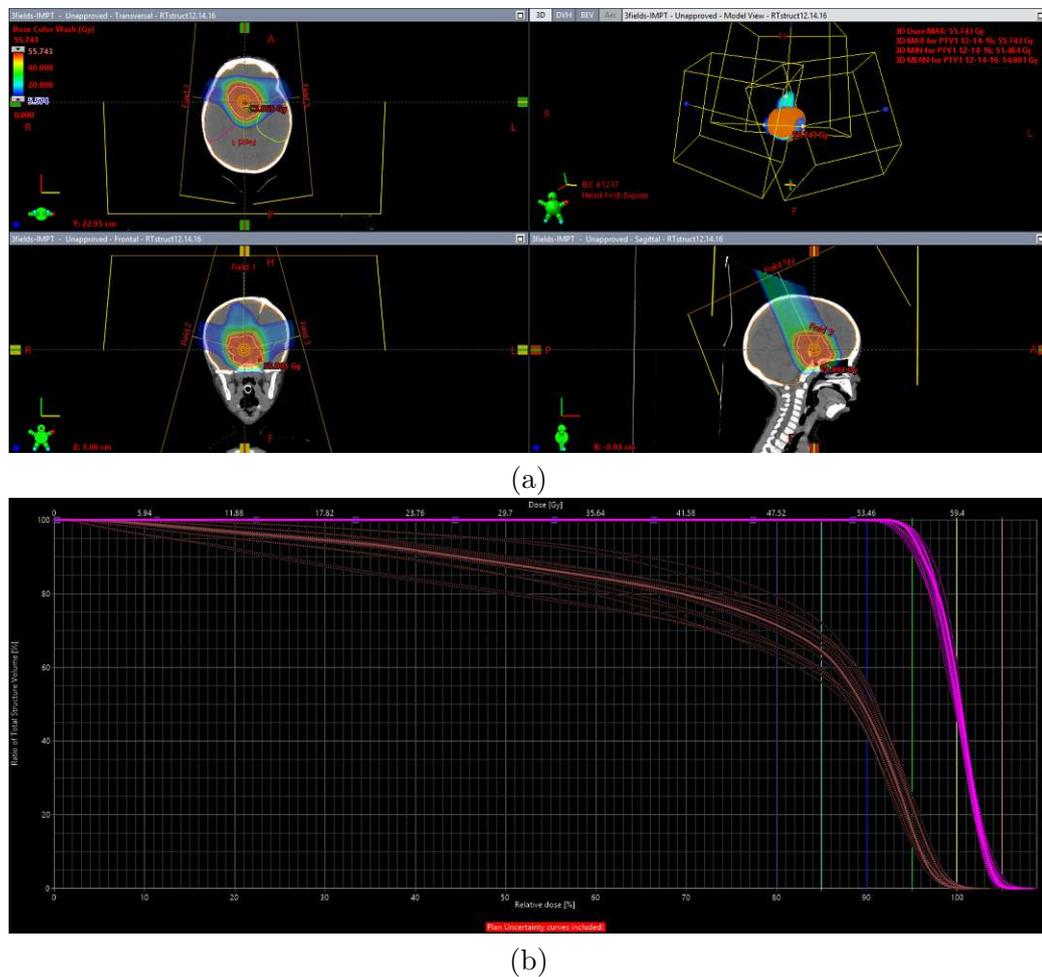


Figure 3.1: Conventional views in available TPS: (a) slice-dose overlay views and (b) DVH, where brown indicates the brain stem (a structure at risk) and magenta indicates the tumor target. The line band corresponds to the plan's robustness. Image produced by Laura Toussaint.

do not allow the user to simultaneously compare and assess multiple plans. They also do not support understanding the involved PT uncertainty factors and how these might affect the treatment outcome. The current workflow leaves a gap for a more efficient sense-making and decision-making process, exploiting the synergy between the human expertise of the researchers or practitioners in PT and the computational power of the TPS.

Guidance techniques [CGM⁺18, CGM19a] could leverage the synergy between domain experts and their systems, but they have yet to be explored within the context of PT or other clinical applications. In clinical applications (and thus also in PT), data exploration through visual interfaces is often complex. There is still a significant lack of *trust* in visual analysis frameworks and *confidence* in the outcomes [GSG⁺21]. It is reflected by the low adoption of visualization solutions in clinical workflows, which indicates that the suitability of visualization frameworks for clinical decision-making scenarios is limited [GSG⁺21]. In PT, uncertainty adds to the sense-making and decision-making complexity [Rai18]. Developing *guidance techniques for uncertainty* within a PT planning visualization system is anticipated to provide researchers and practitioners with a comprehensive view of the planning robustness. Hence, we aim to provide an overview of the entire decision space, which includes all involved uncertainties and their impact on the treatment plan. Our approach is expected to improve the effectiveness and reproducibility of the sense-making and decision-making processes in PT planning.

The *contribution* of this work is the design, development, and assessment of a dashboard as a guided visual interface that enables: (i) the effective comparison of PT plans and (ii) the analysis of the impact of their respective uncertainties. As part of the interface, we investigate and develop *guidance mechanisms* that facilitate navigating through PT plans and their uncertainties in a *multi-level-of-detail manner*, where both the degree of intrusiveness and detail-orientation of the guidance can be tuned and adapted to the needs of the user. Our proposed two-dimensional guidance mechanism, i.e., intrusiveness and detail-orientation, has not been addressed before, while guidance has yet to be investigated extensively in the context of a clinical application. To *evaluate guidance* in PT uncertainty visualization, we propose a framework that focuses on the impact of the guidance on the sense-making and decision-making process.

3.2 Background

3.2.1 Clinical Background

PT requires careful planning to account for several uncertainty factors—from data acquisition to treatment planning and radiation dose delivery. In the planning stage, researchers and practitioners use dedicated TPS, such as Eclipse [SLB⁺18]. They calculate possible treatment plans by deciding on characteristics of the treatment, such as the number of beams, their directions, or specific dose constraints to be fulfilled. Given the computational complexity of the process, only a few, i.e., 2–3, *nominal plans* are computed. Within the few nominal plans, practitioners also consider several uncertainty

factors in deciding which of the plans is the best solution for a patient. Therefore, each treatment plan alternative encompasses a large possible combination of uncertainties [SRM⁺19]. Currently, some uncertainties are accounted for during treatment planning through *robust optimization*.

Despite the significant effort in the visualization community to formalize the definition of uncertainty [RPHL14, GSWS21], our field has yet to adopt a standard definition. For radiotherapy, Raidou defines *uncertainty* as “*any variation in the dose planning outcome, which is produced by an ad-hoc choice or a stochastic process at any step of the radiotherapy pipeline*” [Rai18, p. 14]. This definition outlines the challenging and multi-faceted nature of uncertainty in radiotherapy (and, subsequently, in PT). In common practice, not all sources of uncertainty are addressed at once. Several sources of uncertainty are prioritized due to their high impact on deciding the final treatment plan. For example, previous work focused only on uncertainty due to anatomical variability [FGM⁺20, FMCM⁺21]. Here, we address other types of uncertainty—related to the underlying physics and biological effects of the dose on tissue and their subsequent side effects.

Set-up uncertainties are often investigated in the domain of radiotherapy. They represent the possible deviations in patient positioning on the treatment couch compared to the expected position from which the plan is initially calculated [STS⁺12]. These deviations, in practice, reflect positional variations along the three main anatomical axes of the patient. For example, for each nominal plan, six additional variations along the $\pm x$, $\pm y$, and $\pm z$ axes of the patient are considered. Additional uncertainties are related to the *relative biological effectiveness (RBE)* [Pag18], which is an emerging uncertainty topic in PT that is not yet tackled in TPS. Clinicians use a fixed RBE factor of 1.1 to account for the larger effectiveness of proton- compared to the photon-based RT [MPG⁺17, HDC⁺21]. However, several mathematical models exist that provide a more accurate factor [RFD⁺18]. These calculations also encompass another level of uncertainty with the α/β sensitivity variable per structure, which is essential for the RBE value [KW19, OIT⁺21].

In Figure 3.2, we schematically depict the process resulting in several factors and alternatives for each nominal plan. These need to be considered by clinical physicists to make an informed and complete decision about an optimal plan to follow. Currently, these alternatives are explored and analyzed manually using slice-dose overlay views, DVH plots, or both (Figure 3.1). This solution does not offer a comprehensive view of the different PT plan alternatives’ sense-making and decision-making space. It also does not support clinical physicists in their workflow of investigating PT plan uncertainties to decide on a robust treatment strategy for a given patient.

This work targets decision making by domain experts of various levels of expertise. Thus, the impact of guidance on user confidence and the uncertainty introduced by users’ cognitive biases are different from that expected in the context of domain-agnostic applications. The designed system builds upon the described state-of-the-art TPS, the established workflows in the domain of radiotherapy planning, and the sophisticated scientific models. The target domain experts are familiar with the latter, and thus, they

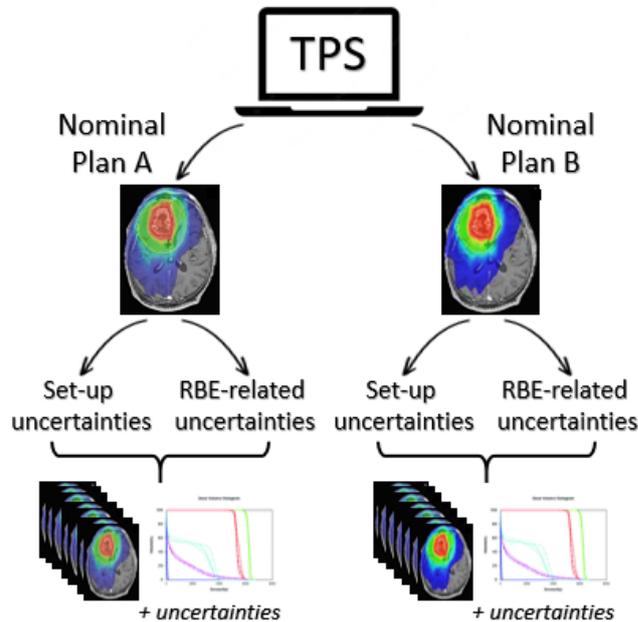


Figure 3.2: Schematic depiction of the uncertainties involved in deciding on an optimal PT plan. For each nominal plan, set-up and RBE-related uncertainties must be accounted for.

have an existing trust relationship with these models. For instance, although they might be aware of the uncertainties they produce, they might not yet be aware of their impact on the simulated therapy plans. This factors into how the users will interact with the system and guidance.

3.2.2 Important Concepts and Definitions

Here, we clarify essential concepts and terms in the chapter:

Clinical goals are clinical guidelines that provide a list of dose exposure limits for each structure based on scientific evidence. This list is used as a reference when physicians inspect dose exposure. Clinical goals help clinicians identify a therapy decision that avoids unwanted damage to structures.

Guidance is ambiguously discussed in visualization and visual analytics (VA), and many different definitions (and interpretations thereof) exist. Guidance comes in many forms: some are prominent, such as text popping up to provide suggestions, and others are subtle, such as visual cues [CGM19a]. In this work, we adopt the definition of Ceneda et al. [CGM⁺17a, p. 1] that defines system *guidance* as “a computer-assisted process

that aims to actively resolve a knowledge gap encountered by users during an interactive visual analytics session”. In our case, the computer-assisted process of analyzing PT plans and their respective uncertainties is facilitated by a two-dimensional mechanism, i.e., intrusiveness vs. detail-orientation, which can be tuned and adapted to users’ needs. This mechanism enables the adaptive use of (visual) cues derived from the data and presented in our interactive visual system to bridge the users’ knowledge gap and inform their analysis. We employ guidance as a means to unveil the impact of uncertainties on the decision-making space and to raise awareness about this impact.

Guidance intrusiveness refers to the degree to which guidance interjects a sense-making and decision-making process. It can be *high* if the guidance amends or supports existing visual plots by, for example, line-styling or *low* when a separate, auxiliary view presents the guidance.

Guidance detail-orientation refers to the detail level, or data resolution, at which guidance is applied. In our case, it can be applied per voxel, per structure of interest, or per slice—depending on the level where the analytical process of the user is occurring.

Knowledge gap indicates the quantifiable difference between the required knowledge to complete an analysis task and the knowledge so far obtained by the user. A gap may arise from different types of knowledge, such as domain knowledge or VA tool knowledge [CAA⁺20]. In the context of our work, the knowledge gap arises from lacking an overview on the entire radiotherapy planning decision space, which is hindered by multi-sourced PT plan uncertainties. Referring to the categorization of Ceneda et al. [CGM⁺17b], the *type* of our knowledge gap relates to an *unknown target* and pertains to the *data domain*, as the workflow revolves around understanding data uncertainties and their impact on the simulated dose data. Based on discussions with domain experts, this knowledge gap cannot only be identified in the interaction with the tool and during the analytical process [CGM⁺17b]—and even before the start of the analytical process. By integrating guidance cues in our tool, we preemptively support the users in their sense-making and decision-making process and provide them with a complete view of the information required to make an informed decision on the optimal plan.

3.3 Task Analysis and Research Questions

After several rounds of discussions among the co-authors of this paper (visualization researchers and medical physicists), we jointly agreed that the current PT workflow is missing two main aspects. First, medical physicists need visualization mechanisms that support them in *comparing and assessing plan alternatives* resulting from the process described in section 3.2 and depicted in Figure 3.2 (**T1**). Second, the workflow should enable the guided exploration of the *impact of different uncertainty types (set-up and RBE) on the final decision* about the optimal plan (**T2**). Making sense of the entire

decision space is a complex process requiring a complete awareness of many uncertainty sources. To unveil how the decision space is affected by uncertainties, guidance is required [CAA⁺20].

Clinical applications often face the problem of *adoptability*. They might be using very complex and unfamiliar visual representations (or interfaces, in general), trying to heavily alter the existing workflow of the domain experts, or both. To counteract this, we provide only simple views that are already familiar to the intended users, and we do not enforce guidance through the analytical workflow.

As our co-designing medical physicists remarked, the workflow should be *tunable to their analytical needs*. The users should be able to decide the degree of aid the system should provide when comparing and assessing plan alternatives and their uncertainties. Tunability, i.e., *control*, is an important requirement for clinical users that empowers them to reach a confident decision on the perceived optimal therapy plan. We use guidance to unveil the impact of uncertainties on the decision-making space and raise awareness about this impact. Nevertheless, the final decision has to be made by the domain expert in a way that is consistent and compatible with the current workflow.

We investigate the following *research questions*:

RQ1 How can guidance assist *PT plan comparison* (**T1**) and *reduce the uncertainty complexity* (**T2**)?

RQ2 Which guidance methods can improve the *confidence* of users in PT plan sense-making and decision-making processes?

RQ3 How can we *validate guidance techniques* designed for uncertainty PT visualization?

To address **RQ1** and **RQ2**, we design a guidance dashboard specific to the scenario of PT uncertainty visualization. It supports the exploration and analysis of multiple PT plans (**T1**) along with their uncertainties and their impact on the decision-making process (**T2**). Only simple representations are employed to support *adoptability*. Specifically for **RQ2**, we investigate different levels of intrusiveness and detail in developing and applying our guidance mechanisms. This approach is necessary to provide the desired level of *control*. Finally, the value of our adopted methods is measured through a domain-expert user evaluation. It builds upon available approaches to account for the guidance impact on the sense-making and decision-making process **RQ3**.

3.4 Design and Implementation

Our designed and implemented dashboard supports comparing PT plan alternatives. The guided navigation of their uncertainties uses a multi-level-of-detail approach with different degrees of intrusiveness. The dashboard is developed in a flexible and extensible

PyCharm environment using Python and Dash. It is the result of an iterative co-design process with clinical physicists, as further discussed in subsection 3.5.1.

3.4.1 Patient Data and PT Uncertainties

We received anonymized planning Computed Tomography (pCT) data with two sets of PT plans for pediatric brain tumor patients. The 3D plans were calculated for illustrative purposes in the context of this work. These included two nominal plans and six different set-up uncertainty scenarios for each nominal plan. In total, we have 14 plan alternatives per patient resulting from the set-up uncertainties. The alternatives are all 3D volumes, where the scalars encode radiation dose in Gray (Gy).

We obtained the respective linear energy transfer (LET) distributions for each plan. These are also 3D scalar volumes, where a scalar encodes the amount of energy an ionizing particle transfers to the material through a unit distance traversal in $\text{keV}/\mu\text{m}$. The data represent the energy deposition density along the depth of the beam, which is an essential variable in estimating damages caused by the radiation dose to the tissues [WWHC11]. The LET is also an input parameter for the different RBE models proposed in the literature.

Moreover, the datasets included a delineation of a total of 93 structures located around/close to the tumor, such as the brain stem or the temporal lobe, and target delineations. We also received a list of maximum-dose clinical goals (maximum radiation dose) as input to inform guidance in calculating the DVHs for the different structures.

In addition to the set-up uncertainties, we deal with RBE-related uncertainties, which must be calculated in real time and integrated into our dashboard. We calculate the RBE values using the LET distribution and eight known models in the field [RFD⁺18]. Thus, we obtain eight RBE calculations for each plan, which may differ from each other. Additionally, we use a default structure sensitivity value α/β of 2.0 Gy, but the interface enables the user to control these values in real time while comparing PT plans. This functionality enables the user to unravel the RBE uncertainty further.

3.4.2 Uncertainty Guidance Design

We propose a *two-dimensional formulation of uncertainty guidance*, which targets complex scenarios where a multi-level-of-detail —yet controlled— exploration of uncertainties is required. Table 3.1 shows an overview and categorization of the guidance mechanisms developed as part of our dashboard and further discussed in subsection 3.4.3.

The first dimension represents *intrusiveness*. As discussed in subsection 3.2.2, *low-degree intrusiveness* does not “intrude” the user’s analytical visual space; instead, it employs on-demand, auxiliary solutions. *High-degree intrusiveness* becomes intrinsic to the user’s analytical visual space by altering it. Our definition indicates that highly intrusive guidance is displayed within the user’s direct perceptual focus or is integrated into conventional views, such as *new encodings within* the slice-dose overlay or the DVH view.

Table 3.1: Guidance overview and categorization using two dimensions (intrusiveness and detail-orientation) and employed in our dashboard to compare the available PT plans and navigate through their respective uncertainties.

		Detail-Orientation		
		Per Voxel	Per Structure	Per Slice
Intrusiveness	Low	Slice–dose overlay (Figure 3.3) & Explicit comparison view (Figure 3.4)	RBE uncertainty violin plot (Figure 3.6)	Uncertainty indicator (Figure 3.8)
	High	On-demand uncertainty distribution plot (Figure 3.5)	Targeted stylization of DVH lines for structures of interest (Figure 3.7)	Re-adjustment of the presented dose based on selected RBE models

Oppositely, less intrusive guidance comes within separate, supplementary views that can be added on-demand. This salient encoding is intended to draw the user’s attention proactively. Intrusiveness relates—but is not entirely equivalent—to the controllability characteristic of guidance, as discussed by Ceneda et al. [CAA⁺20]. The system allows the user to control the degree of guidance and, subsequently, the degree of intrusiveness to comply with their analytical process. This feature also resembles the *refine* stage, as described by Sperrle et al. [SJB⁺21]. As a difference, our approach does not include automated control; instead, it leaves tuning to the user.

The second dimension represents *detail-orientation*, which relates to the level of detail, i.e., voxel, structure, or slice level, at which the guidance works. At each level, we investigate and compare the impact of uncertainties on the patient plan. The structure level is the most common, as planning often considers entire organs [SRM⁺19]. The voxel level provides more granular detail on sub-parts of the organ where significant differences might occur. The slice level is an aggregation essential for efficiently comparing doses and identifying issues at very impactful slice ranges, for example, close to tumors or structures.

In designing our guidance mechanisms, we consider additional points. First, we integrate uncertainty guidance with anatomical and DVH plots, maintaining *conventions* in the PT domain to promote adoptability. Second, we employ the guidance plots as *abstract visual cues*. We abstract the plots if low-level details are not meaningful for the analytical process of the user. For example, if we are interested in a high-level comparison of two or more dose distributions, we largely omit axis labels and ticks as unnecessary detail. This way, we accommodate many plans and uncertainty factors in an abstract comparative view to provide indications of plan differences that guide the user through the decision-making

process. Finally, our guidance techniques encompass *visual organization techniques*. Thus, we do not introduce any new plots. We only rearrange and regroup existing ones. The following subsection discusses how we implemented the guidance-enriched PT planning dashboard.

3.4.3 Dashboard Implementation

We followed a detailed–context–overview approach while building our dashboard [LBS⁺19]. The final dashboard evolved through an iterative co-design process with domain experts (further described in section 3.5). It resulted in the inclusion of different views that enable the users to explore and compare details and summaries of plan-dose overlays with their uncertainties. The dashboard also provides several control features to enable dynamic guidance based on user interaction.

As *input*, the dashboard requires DICOM (Digital Imaging and Communications in Medicine) data that comprise a patient’s pCT scan slices, information about the structures of interest (delineations and dose-limitation clinical goals), and the respective dose and LET distributions. As *output*, the dashboard displays a comparative view of different PT plans and their uncertainties (RBE and set-up). Concerning the *available knowledge*, the domain experts are knowledgeable and experienced in reading and comprehending dose/LET plan visualizations overlaid on pCT scans together with structure contours and DVH plots (Figure 3.1). The domain experts are also knowledgeable about the nature of uncertainties. However, they could benefit from visualizations that support them in understanding and comparing uncertainties from different plan alternatives (**T1,2**). In this case, conventional plan visualizations and DVH plots are insufficient, as discussed in section 3.2. This is, thus, the *knowledge gap* we attempt to bridge.

Slice–Dose Overlay View Figure 3.3 includes the typical slice–dose overlay view on the pCT slices along the three main anatomical axes for two plans. This slice-based view is conventionally used in PT to represent the dose distribution on the patient’s anatomy at each voxel position [SPBR20]. In the background, the pCT scan is presented using a grayscale, and the dose- (or LET-) plan is overlaid using a rainbow color map. Additionally, structure contours are overlaid with distinct colors. The rainbow color map and the contour colors are retained to follow standard practices and conventions in PT visualization [SPBR20]. This view supports **T1**. To optimize space on the interface for the comparison, we position the three anatomical planes vertically, as opposed to the configuration in Figure 3.1 (a). Our collaborators did not deem a 3D view necessary.

To compare two plans, the domain experts currently must exchange views by moving back and forth between plans. Plumlee and Ware [PW06] suggest multiple windows to compare complex data that the users cannot easily hold in their visual working memory. In our dashboard, we adopt a juxtaposition approach where the two plans are depicted side-by-side. This supports plan comparison through a simultaneous and linked exploration of the slices. It is a *low-intrusiveness* and *per-voxel* guidance mechanism, which allows a one-by-one comparison of two dose- (or LET-) plans for one patient at a

voxel level. Superposition is not an adequate choice, as the underlying planning CT scan has to be retained for anatomical context.

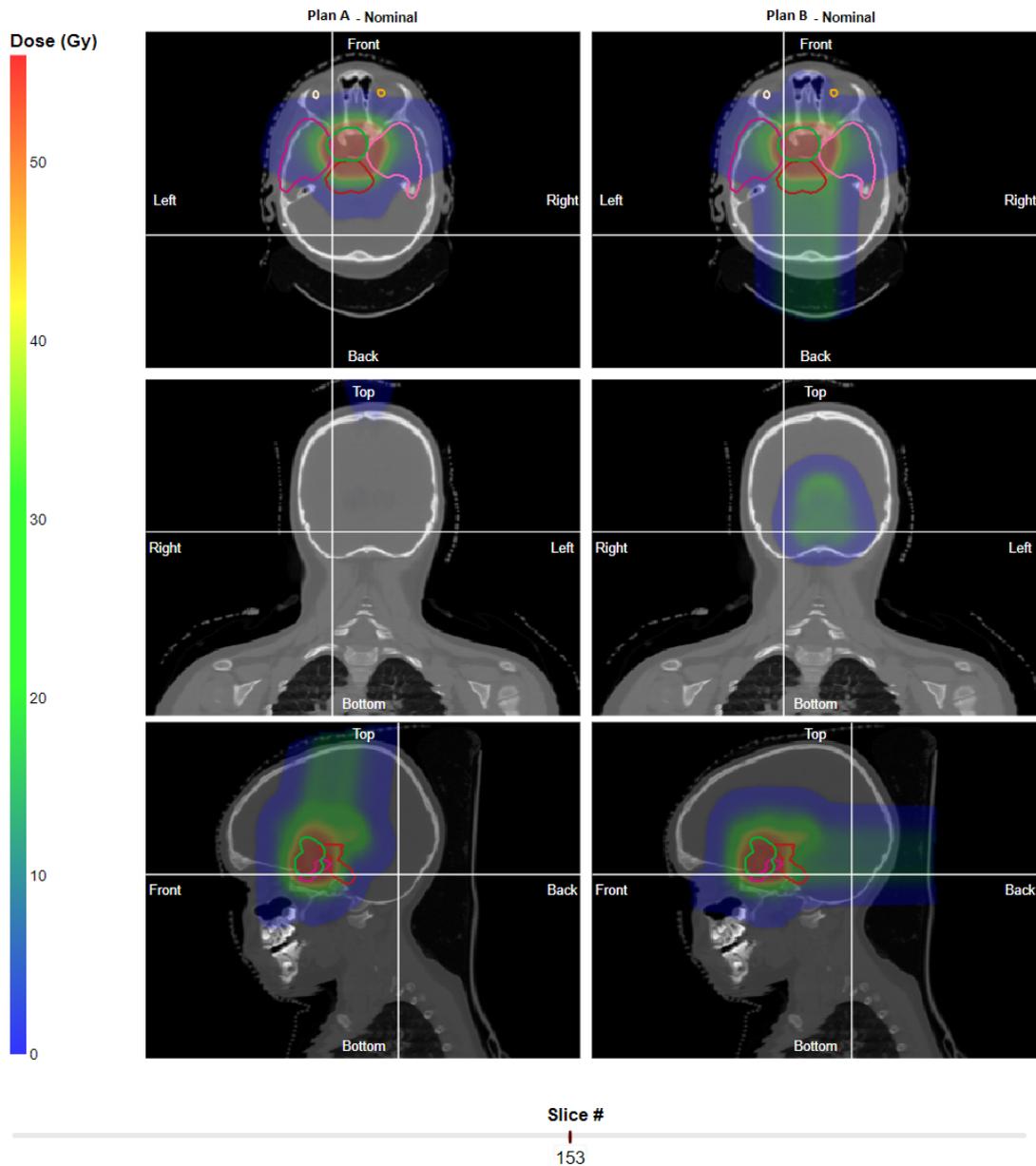


Figure 3.3: pCT slices overlaid with dose distributions and structure contours from two plans (in the two columns) on the three main anatomical planes (in the three rows) of the patient for one slice (Slice 153). The contour colors (colormap to the left) follow the conventions in the clinical environment.

Explicit Comparison View For a more explicit comparison of two plans, we use the view shown in Figure 3.4. Explicit encoding of the computed difference provides better precision of the image difference [KPBG13], which supports the comprehensibility demand of a guidance encoding. The view encodes the difference between the primary (top) plane views in Figure 3.3. It includes the typical anatomical slice with a semitransparent dose overlay that results from the absolute difference between the dose or LET values of the first and the second plan. A calculated value may be positive or negative, based on which plan receives a higher dose at the specific voxel position. The difference is encoded in a red-to-blue diverging color scale. Red represents a higher dose value for the first plan, and blue represents a higher dose value for the second at a given voxel position. The view can additionally include structure delineations, as used conventionally in the domain of PT planning. It implements a *low-intrusiveness* and *per-voxel* guidance, which supplements the slice-dose overlay view. The explicit comparison view supports **T1**.

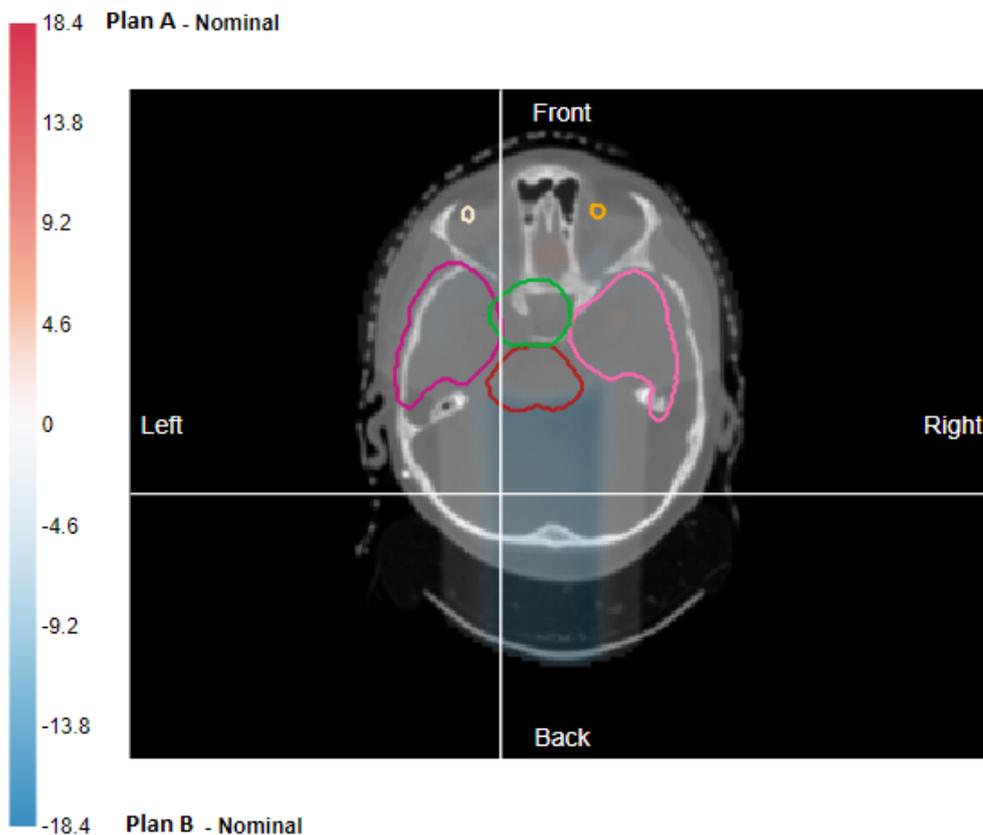


Figure 3.4: The explicit comparison view encodes the differences between the two plans of Figure 3.3 using a divergent red-to-blue colormap to indicate different doses between the two alternatives at each voxel. The contour colors follow the conventions in the clinical environment.

Uncertainty Distribution Plot The user can hover over a voxel position in an ROI that receives a radiation dose, for example, a voxel within the brain stem. In this case, a plot of the dose distributions of the alternative plans at that specific voxel with accompanying uncertainties will appear. Several comparison techniques are possible. Due to the high-intrusiveness nature of this encoding, we use superposition to provide a direct visualization of the trade-off between the two plans [KPBG13].

The plot compares at a high level and highlights differences in the distribution of all possible dose values at a voxel. It is based on all different RBE and set-up uncertainty factors for each plan and structure (Figure 3.5), and supporting **T2**. We depict each plan with a distinct color (red or blue, following the explicit comparison convention). The rugplot [Tuf01] under the horizontal axis further enhances this representation by visualizing the marginal distribution of the data as marks along the axis. Rugplots are often coupled with distribution plots to enhance the view on the raw data used in plotting the distribution. This guidance mechanism provides users with a many-to-many robustness indication in comparing two plans *at a voxel level* with a *high degree of intrusiveness*.

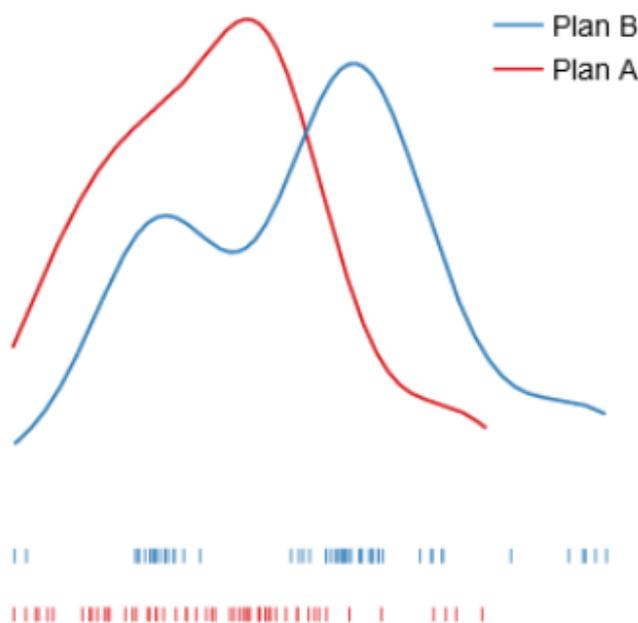


Figure 3.5: The distribution plot pops up when hovering over a voxel position within a delineated structure receiving a radiation dose. It reveals the high-level uncertainty distributions of the two plans, indicated with the two different colors. The rugplot, i.e., the marks on the horizontal axis at the bottom, further enhances the abstract comparative representation.

RBE Uncertainty Violin Plot The view in Figure 3.6 provides a guidance of *low degree of intrusiveness* to the user on the robustness of plans. Robustness is based on the uncertainty distribution introduced by the calculated RBE factor of *each structure of interest* (**T2**). We represent the RBE uncertainty distribution for each structure of interest with a split violin plot, as seen in Figure 3.6. It displays at a high level the distribution of the possible values based on calculating RBE at each scenario of the plan with each of the chosen calculation models and other parametrizations for the structures of interest.

We considered alternatives like common box plots or sparse representations, such as those proposed by Wentzel et al. [WHL⁺19]. Although box plots might be more familiar to domain experts, violin plots are advantageous for showing the entire data distribution [HN98]. They also facilitate the identification of differences across distributions. This characteristic enables the users to detect marginal trade-offs between the compared plans, which is crucial in our scenario. The user can zoom into areas of interest. Additional statistical annotations are provided on-demand to support the comparison of the two distributions—showing also exact numerical values. Eight models are used to calculate the RBE factors from each set-up scenario at each voxel position (in total, 56 different alternatives) for each structure and both plans. In Figure 3.6, a subset thereof is shown.

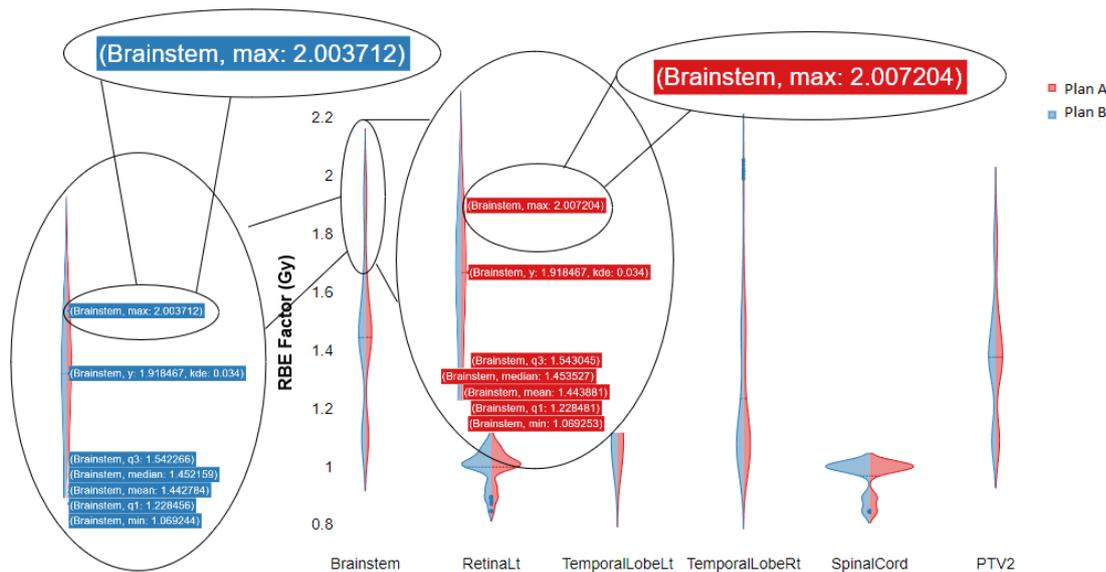


Figure 3.6: Comparative uncertainty distribution of possible RBE factors for the structures of interest. The two plans are shown with two distinct colors. The RBE uncertainty calculations include many factors, such as different RBE models and structure sensitivities (α/β ratio). When hovering, the user is provided with additional annotations to support the comparison. The calculations have been done for six structures of interest.

Targeted Stylization of Dose Volume Histogram (DVH) Lines The DVH plot is a typical depiction in the RT domain that displays the distribution of the dose over volume percentages of structures (see Figure 3.1(b)). The distribution of dose values to volume percentages is calculated from the DICOM data files, which contain the delineations of the structures of interest. In the DVH plot, users can inspect the dose or LET volume distribution. DVH plots accommodate a many-to-many *per-structure* comparison (T1,2). Different plans and uncertainty factors can be computed and plotted with color coding and different line styles. In Figure 3.7, we illustrate two plans in the same DVH plot with several structures of interest. The user selects and deselects structures for each plan through the active legend. Line stylization supports an *intrusive* guiding strategy. A distribution line is drawn with a larger width if the corresponding structure received a maximum dose that exceeds the clinical goals. In Figure 3.7, this occurs for the temporal lobes in plans A and B. Adjusting the line width provides a more expressive visual guidance cue, given the lightness of the conventional colors used to represent the structures. A less intrusive mechanism may involve uncertainty bands [ZKL17], where the per-structure set-up uncertainty is displayed as a band around the line.

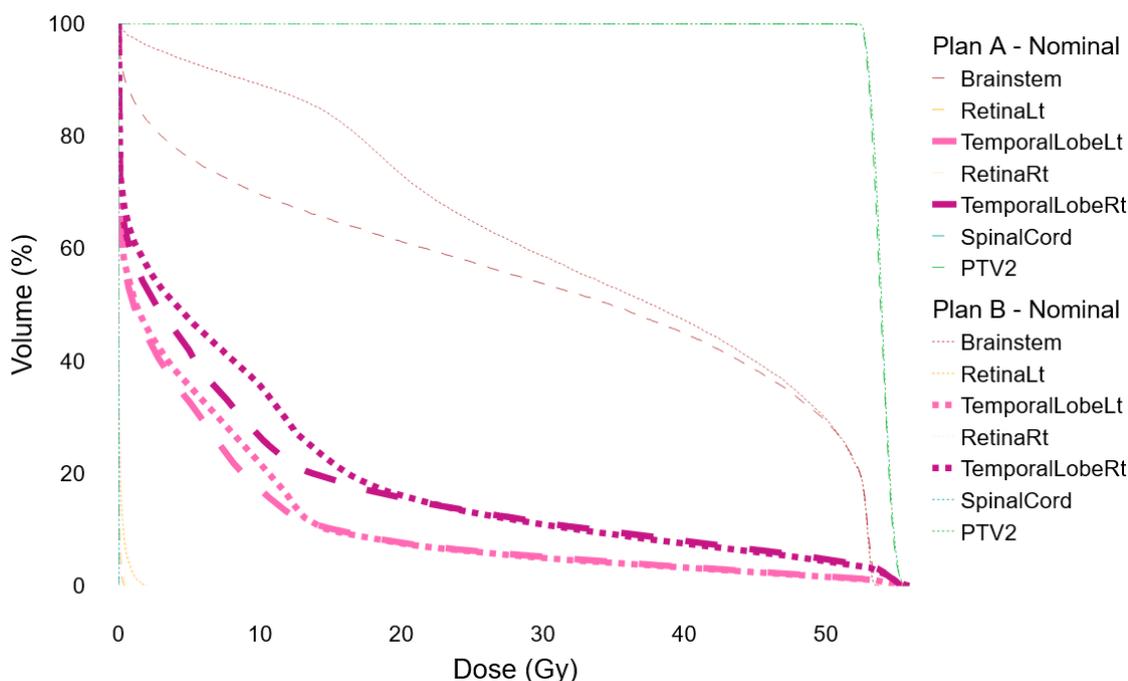


Figure 3.7: The Dose Volume Histogram (DVH) for the structures involved in two plans (plan A indicated with dashed lines and plan B with dotted lines). Line stylization guides the user in identifying structures at risk (encoded by the increased line width, for example, for the temporal lobes in plans A and B). The colors have been selected based on TPS conventions.

Uncertainty Indicator A heatmap plot summarizes the plan robustness per slice, i.e., which plan has less uncertainty for each slice. This view supports **T2**. We use a grayscale heatmap to avoid user confusion, given that the dashboard already employs other color encodings dictated by field conventions.

We calculate the average dose exposure and standard deviation for each slice, which resulted from the different RBE calculation models. The standard deviation σ indicates the variability in the uncertainties throughout all possible alternatives. σ is encoded in a grayscale where dark grey indicates a high σ , i.e., a slice with high uncertainty. An example of the uncertainty indicator view is shown in Figure 3.8, juxtaposed are two plans and the different slices of the volumes. In this example, slices beyond number 230 have high σ s, i.e., uncertainty, in plan A and low σ s in plan B. Brushing and linking allow the user to select a particular slice by directly clicking on the uncertainty indicator to investigate the slice further. The view integrates a *low-intrusiveness guidance* to support uncertainty plan comparison at a *slice level*.

Re-adjustment of the Presented Dose based on Selected RBE Models The user can select which models to include in calculating the average RBE factor per voxel. It is done through several controls, further described below. Changing the model will trigger a real-time recalculation and adjustment of the views affected by changes in the RBE computation. It can be experienced as a *high-intrusiveness solution*, mainly applicable to *slice-based* exploration and analysis. The re-adjustment mechanism supports **T1** and **T2**.

Controls The dashboard is supported through several controls that accompany the guidance strategies. Some controls are shared between views, and others are restricted to an individual view. In total, the dashboard includes five controls, one for: manipulating the slice-based view, the inclusion/exclusion of structures of interest, setting up the clinical goals and sensitivity value for each structure (Figure 3.9 (a)), exchanging the RBE models and set-up scenarios (Figure 3.9 (b)), selecting the level of guidance intrusiveness that the user prefers (Figure 3.9 (c)).

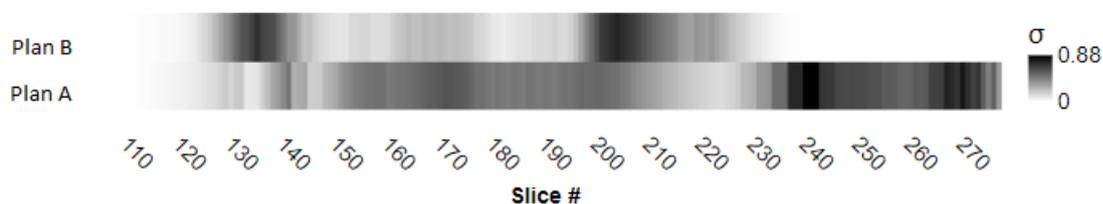


Figure 3.8: The uncertainty indicator view provides a per-slice indication of the uncertainty for two plans and all slices. A darker gray value indicates a higher σ , i.e., higher uncertainty, in the given slice.

3. UNCERTAINTY GUIDANCE IN PROTON THERAPY PLANNING VISUALIZATION

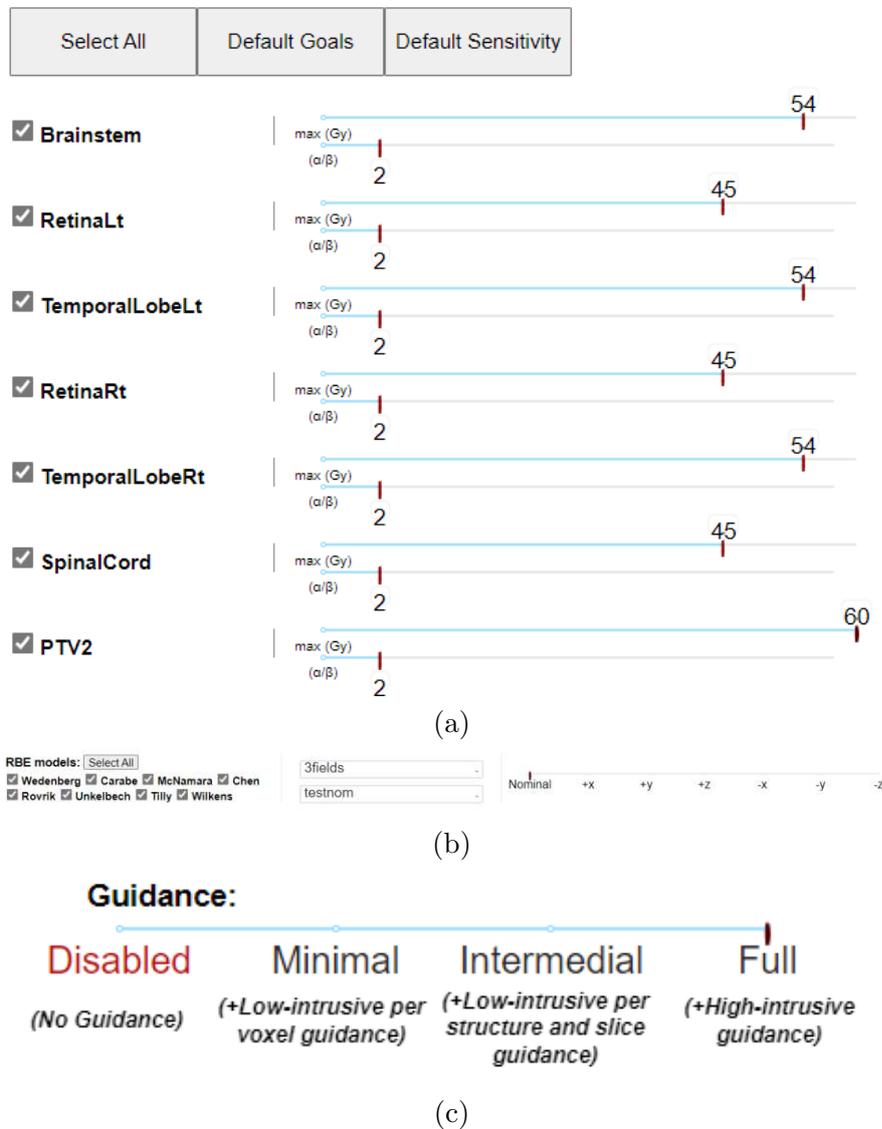


Figure 3.9: Three of the five controls accompanying the guidance mechanisms of our dashboard: (a) provides a check-list of the structures of interest and sliders for setting their clinical goals and sensitivity values, (b) provides a check-list of the included RBE models, a drop-down menu to choose plans for comparison, and a slider for selecting set-up uncertainty scenarios, and (c) provides an interface to select the level of guidance intrusiveness.

Making a selection through one of the controls has a linked impact on one or more views of the dashboard. For example, the controls of Figure 3.9 (a) provide an input mechanism for the user to include or exclude structures and set their sensitivity values and clinical goals. Modifications directly impact dose calculation, which is then reflected on the slice–dose overlay view, the explicit comparison view, and, subsequently, on all the other views of the dashboard. Similarly, the controls of Figure 3.9 (b) concern the inclusion of different RBE uncertainty sources and set-up uncertainties, affecting all possible dashboard views.

Finally, the slider in Figure 3.9 (c) controls the degree of guidance intrusiveness in the dashboard views. The user can decide how intrusive the uncertainty guidance should be in her sense-making and decision-making process. The control starts from disabled guidance, where the conventional slice-based views in Figure 3.3 and the DVH plot in Figure 3.7 are presented. In the minimal and intermediate degrees of intrusiveness, low-intrusive guidance is added to the interface. In the former, the explicit comparison view (Figure 3.4) is included, and in the latter, the RBE uncertainty violin plot (Figure 3.6) and uncertainty indicator (Figure 3.8) are added. The full guidance mode includes the intrusive guidance that styles lines in the DVH plot (Figure 3.7) and the uncertainty distribution plot on slice pixel hovering (Figure 3.3). It also presents the RBE calculation controls in Figure 3.9 (a) and (b).

3.4.4 Usage Scenario

In this section, we present a usage scenario to showcase the full functionality of our dashboard. We expect the users to start their navigation with the lowest degree of guidance, as this coincides with current practices. Initially, the slider under the panels is used to browse the linked anatomical slices in Figure 3.3. The users scroll through the data until they reach a slice that displays the region(s) of interest (ROI). For example, in Figure 3.3, these could be the temporal lobes (indicated in pink and magenta), the brain stem (indicated in red), and the tumor volume (indicated in green). They can, subsequently, add more guidance and consult with the explicit comparison view in Figure 3.4. Here, they can identify a significant difference in dose radiation in specific ROIs (**T1**), i.e., the posterior part of the brain receives a higher radiation dose with plan B (indicated by the blue color). In Figure 3.9 (a), the users can deselect the contours of the least important regions. Hovering over a structure allows them to explore the uncertainty in the voxel values, for example, in Figure 3.5, which shows a shift between the two plan distributions (**T2**). To receive a more comprehensive picture, they can change the displayed data or select a higher dose value range to display. Users inspect the plot in Figure 3.8 to identify particular slices where a significant uncertainty variance occurs. In this example, they can select slice 138, where plan A exhibits a significant uncertainty variance indicated with a darker color. Slice 138 is also depicted in Figure 3.4, where a difference in dose is noticeable.

To complement their findings, the users can opt for the highest degree of guidance and request to additionally explore the DVH plot in Figure 3.7. In this example, the increased

line thickness indicates structures that may be at risk (**T1,2**). For instance, the right temporal lobe indicated with the dashed, thick, magenta line might be a structure at risk in both plans A and B. In contrast to this, the retina (light pink line) is not a structure at risk in either of the two plans. The users can then customize the clinical goal for an interesting structure at risk and deselect the rest. They navigate the differences between the dose distributions in the structure in both plans using the plot shown in Figure 3.5. Here, the two distributions differ by a shift, indicating that plan B administers a higher radiation dose overall. To get a clearer picture of the plan’s robustness at a structure level, users can further explore the violin plots in Figure 3.6. The plots may indicate significant differences between the plans for each of the structures. In this example, the differences are slight, as the two sides of the violin plot are almost identical. Finally, users can also customize the sensitivity values and RBE calculation models in Figure 3.9 (a) and (b), respectively, and include the sensitivity uncertainties in the analysis. Now, users can make decisions relying on a high or low degree of guidance intrusiveness in our tool.

3.5 Evaluation and Results

We followed an iterative co-design process with domain experts, where the views and user interactions discussed in section 3.4 were developed in tight collaboration with domain experts. This approach has facilitated understanding problems in the PT planning domain [MCJ22]. The iterative design included *three formative sessions* to inform our design decisions (subsection 3.5.1). After the finalization of the design, we conducted a *user evaluation* (subsection 3.5.2). The aim was to evaluate whether the dashboard, with its guidance mechanisms at different degrees of intrusiveness, influences the perceived confidence and supports the users in their sense-making and decision-making process. We included a total of six domain experts (Table 3.2) throughout the formative sessions and the user evaluation (Table 3.3).

Table 3.2: Domain experts participating in the evaluation sessions and their experience (in years) in RT research and practice.

	Occupation	Research years	Practice years
P1	Medical Physicist	25	5
P2	PostDoc Researcher	5	0
P3	Medical Physicist	19	24
P4	PhD Student	1	0
P5	Medical Doctor	6	3
P6	Medical Physicist	18	30

Table 3.3: Participation of domain experts in the different stages of the evaluation (formative sessions vs. user evaluation).

	Evaluation Stage			
	Formative Sessions			User Evaluation
	1	2	3	
P1	✓	✓	✓	
P2	✓	✓	✓	✓
P3	✓			✓
P4				✓
P5				✓
P6				✓

3.5.1 Formative Evaluation

We conducted three formative evaluation sessions to inform our design decisions. As seen in Table 3.3, *P1* and *P2* were involved in the entire design process. They initiated this project by providing data, initial information regarding current practices and the clinical workflow, as well as ideas for improving the current sense-making and decision-making process.

The *first formative session* was conducted together with *P1*, *P2*, and *P3* in a joint session. We discussed in this session a low-fidelity prototype sketch (Figure 3.10) to elaborate further on how the users imagine the design of the dashboard. We discussed *which tasks users would like to be able to perform* and *how their current workflow could be improved*. We also dealt with *visual cues* and how we could design them better to ensure *compliance* with field conventions. *P1* and *P2* highlighted the unresolved visualization of uncertainties in the process as their current knowledge gap. Presently, the field of PT is discussing the possibility of selecting one or two RBE calculation models instead of the 1.1 RBE factor, which is the current clinical default value. After this session, we improved the comparison encoding in the RBE uncertainty violin plot based on discussions with the expert collaborators. We also condensed the slice views to provide a permanent display of the three plane views—axial, sagittal, and coronal. Moreover, we included domain-familiar visualizations, i.e., the DVH plot and slice views. Finally, we added controls to enable user customization and a heatmap plot to indicate plan robustness. It resulted in the second design (Figure 3.11).

The *second formative session* was conducted individually with *P1* and *P2* to discuss the second design iteration. This session was targeted toward *concretizing the design of the individual views*. We started by discussing the views and functionalities of the dashboard. Then, *P1* and *P2* obtained remote control and freely explored the dashboard. We received valuable feedback regarding functionality and visual cues to inform our third design iteration. After this session, we changed the parallel coordinates plot to a violin plot to represent more accurately the RBE uncertainty. We rearranged the views based on feedback about the anticipated analysis process. Furthermore, we made the slice views

3. UNCERTAINTY GUIDANCE IN PROTON THERAPY PLANNING VISUALIZATION

larger and permanently displayed the three slice views of both plans in a juxtaposed manner to better support comparisons. Finally, we redesigned the heatmap plot into an aggregated version. It resulted in the third design (Figure 3.12).

In the *third formative session*, we discussed the improved iteration of our dashboard. The purpose of this session was to *receive final feedback from both co-designers* before putting the dashboard out for the main user evaluation with a larger group, which is discussed

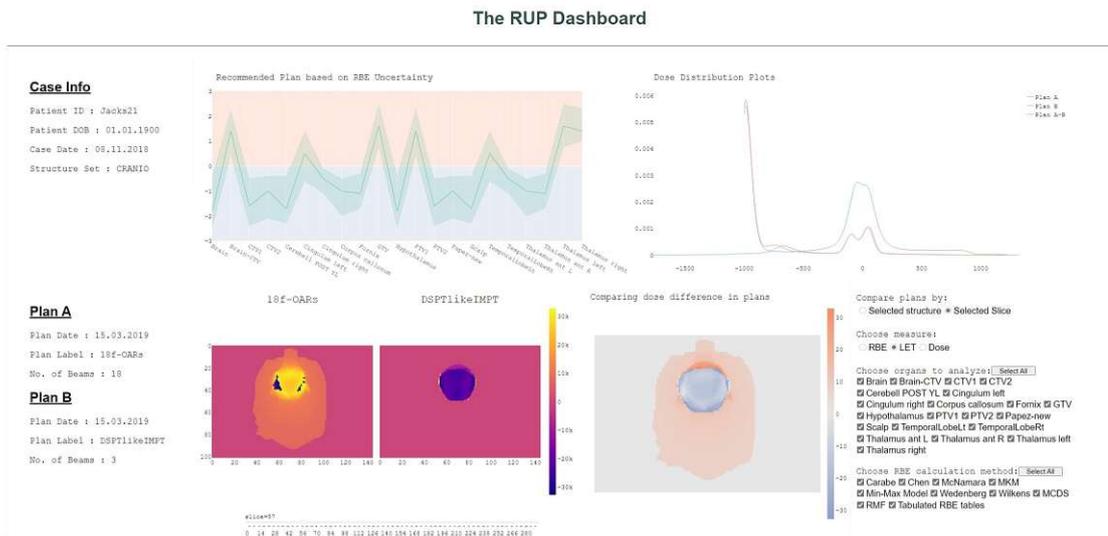


Figure 3.10: Low-fidelity prototype used in the first session.

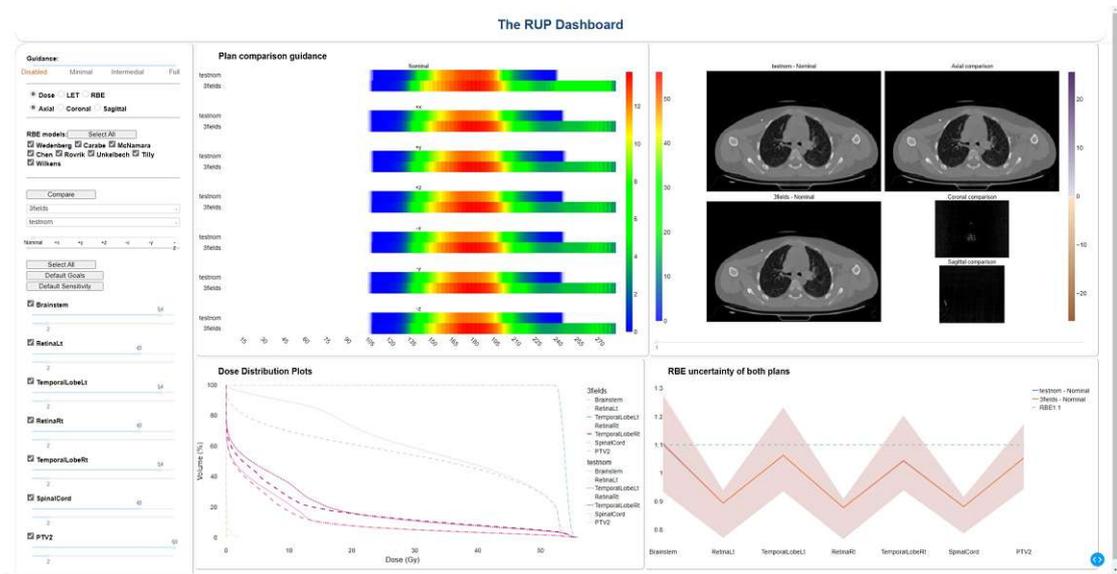


Figure 3.11: Interactive improved prototype used in the second session.

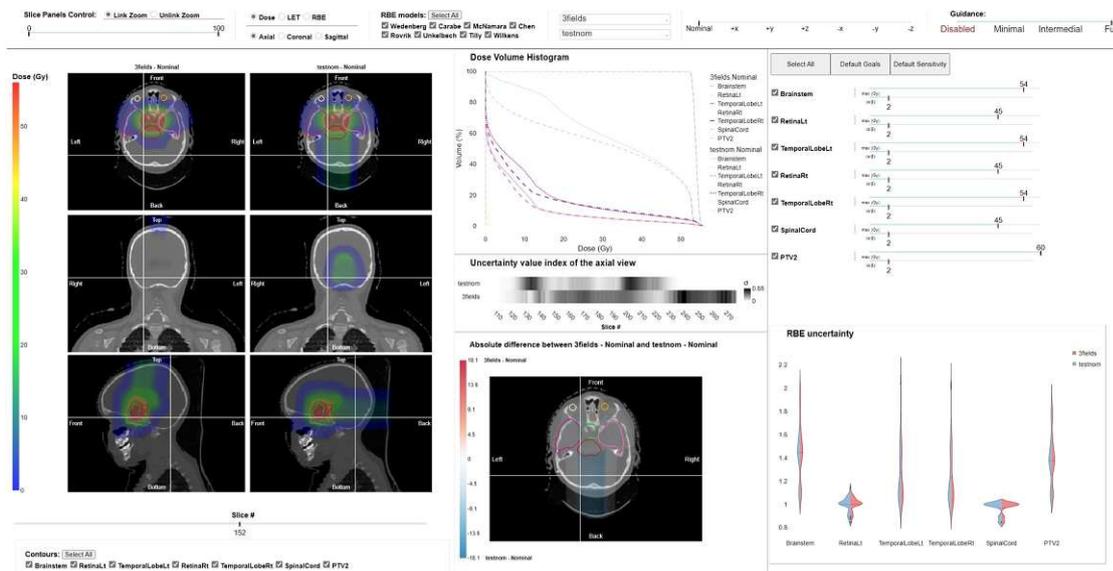


Figure 3.12: Final dashboard employed in the user evaluation.

in the subsection 3.5.2. As illustrated in Figure 3.10, Figure 3.11, and Figure 3.12, the dashboard evolved through several stages of an iterative process, where together with domain experts we identified design choices appropriate to fulfill their tasks. Some views were included in the first two versions of the dashboard, such as different comparative views or parallel coordinates to show uncertainties, but did not make it to the final dashboard because they were not considered effective or insightful enough.

3.5.2 User Evaluation

User Evaluation Design In our study, we evaluate the impact of guidance techniques on the sense-making and decision-making process. We also assess the potential of the uncertainty visualization approaches in reducing the perceived complexity in the calculated PT plans. Specifically, we aimed to answer the following questions:

- EQ1** How does the user employ the dashboard with regard to the given *uncertainty guidance mechanisms at different levels of intrusiveness* (low vs. high)?
- EQ2** How does the user employ the dashboard with regard to the given *encodings at different detail-orientations* (voxel, structure, and slice)?

To answer the questions, the user evaluation design was informed by the taxonomy of scenarios presented by Lam et al. [LBI⁺12]. We evaluate first the *visual data analysis and reasoning (VDAR) scenario*. The study participants explored real-patient data in a controlled environment, while we observed their behavior while employing the dashboard. Moreover, we wanted to assess how the guidance mechanisms affected *user*

3. UNCERTAINTY GUIDANCE IN PROTON THERAPY PLANNING VISUALIZATION

Table 3.4: The 7-point Likert scale questionnaire of the ICE+SUS (inspired by previous works [WAM⁺19, Bro96]), completed by the study participants.

Questionnaire

A. What is your recommended radiotherapy plan? How did the visualization system inform your decision?

B. Please mark the box that best reflects your reaction.

<i>Insight</i>	Strongly Disagree						Strongly Agree
1. The visualization helps me detect and understand the differences between the two plans	<input type="checkbox"/>						
2. The comparison between plans is difficult	<input type="checkbox"/>						
3. The visualization provides different level-of-detail information when exploring the data	<input type="checkbox"/>						
4. The information provided by the panels in the dashboard is redundant	<input type="checkbox"/>						
<i>Confidence</i>							
5. The visualization reduces the complexity of the data	<input type="checkbox"/>						
6. The visualization uses meaningful and accurate visual representations	<input type="checkbox"/>						
7. The visualization is a misleading representation of the data	<input type="checkbox"/>						
8. The knowledge presented on the screen is complicated	<input type="checkbox"/>						
<i>Essence</i>							
9. The system does not provide more value than existing tools	<input type="checkbox"/>						
10. The visualization helps me make a decision about a plan recommendation	<input type="checkbox"/>						
11. The visualization provides an understanding of the uncertainties in the data	<input type="checkbox"/>						
12. The visual representation of uncertainty in the data did not impact my decision	<input type="checkbox"/>						

performance (UP) in exploring the data and making a decision. For this, we used an evaluation framework, which we called ICE+SUS, as it was inspired by both the ICE-T questionnaire by Wall et al. [WAM⁺19] and the System Usability Study (SUS) popularized by John Brooke [Bro96]. In this questionnaire (Table 3.4), six statements were worded negatively, and the remaining six were positive statements to minimize acquiescence and extreme response biases. The participants were requested to fill out this questionnaire after using the dashboard with different guidance levels. Furthermore, we evaluated *user experience (UE)* by assessing whether the encodings effectively reduced perceived uncertainties and increased the users' confidence. For this, the responses to the ICE+SUS survey and answers to additional open-ended questions were solicited. The open-ended questions regard what could be improved, reworked, or seen as useful among the uncertainty guidance mechanisms of the dashboard.

User Evaluation Course The final dashboard, as resulting from the formative sessions (Figure 3.12), underwent an online evaluation with five participants (*P2–P6*). *P1* did not participate due to extensive familiarity with the dashboard. We conducted individual online and recorded sessions using screen sharing and remote control. We spent the first few minutes explaining the views of the interface and basic functionalities of the dashboard until these were clear to each participant. Then, the participant was given control to freely explore the dashboard without our interference.

The user evaluation was divided into three sub-sessions, where the intrusiveness level of the guidance mechanisms was increasing. To control for bias, we randomized the order of the sub-sessions. At the end of each sub-session, the users made a recommendation on

the plans and filled out the questionnaire (ICE+SUS and the open-ended questions) to address the UE and UP scenarios. The results of this part are discussed in the remainder of this section. The VDAR part resulted in use cases; one example is described in subsection 3.5.3.

User Evaluation Outcomes The results of the study with regard to the intrusiveness dimension (**EQ1**) are presented in Table 3.5. They show a constant improvement with increasing intrusiveness in the guidance mechanisms, in terms of *gained insights* and *essence*, i.e., added value to the current workflow. This improvement was constant

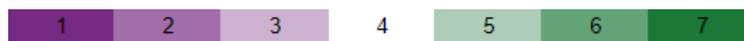
Table 3.5: ICE+SUS feedback results from the three sub-sessions of the user evaluation: with full guidance (high intrusiveness), with intermediate guidance, and without guidance (low intrusiveness). Point 1 indicates the strongest disagreement, and 7 denotes the strongest agreement.

Full-guidance				
	Insight	Confidence	Essence	Average
Participant 2	6.00	5.00	4.75	5.25
Participant 3	6.25	4.75	5.75	5.58
Participant 4	5.25	4.50	4.75	4.83
Participant 5	4.50	4.25	4.50	4.42
Participant 6	6.00	3.25	4.25	4.50
95% C.I.	5,60 ± 0,89	4,35 ± 0,84	4,80 ± 0,71	4,92 ± 0,62

Intermediate Guidance				
	Insight	Confidence	Essence	Average
Participant 2	5.75	5.25	4.75	5.25
Participant 3	4.25	4.75	5.75	4.92
Participant 4	4.25	4.50	6.00	4.92
Participant 5	4.50	2.75	3.25	3.50
Participant 6	5.75	3.75	4.00	4.50
95% C.I.	4,90 ± 0,97	4,20 ± 1,21	4,75 ± 1,44	4,62 ± 0,84

No Guidance				
	Insight	Confidence	Essence	Average
Participant 2	4.25	4.50	3.25	4.00
Participant 3	2.25	6.00	5.00	4.42
Participant 4	5.50	5.00	4.50	5.00
Participant 5	4.75	2.75	2.50	3.33
Participant 6	4.25	4.75	3.50	4.17
95% C.I.	4,20 ± 1,49	4,60 ± 1,47	3,75 ± 1,24	4,18 ± 0,75

Legend:



across all participants, except for P_4 , who provided the most positive feedback for low intrusiveness approaches. A reason might be in P_4 being the least experienced among our study participants (see Table 3.2). On the other side, participants with the highest RT research and practice experiences, i.e., P_3 and P_6 , commented that they gained the least insight with the no-guidance dashboard. Their insights increased when presented with full guidance. However, we noticed a decrease in *reported confidence* as more intrusive guidance was introduced to the interface. The two highly-experienced participants (P_3 and P_6 , Table 3.2) also expressed less confidence in the uncertainty encoding. The results may imply that, in the current state, the dashboard does not increase the perceived confidence of the users. However, it provides them with significantly more insight through the fully guided interface and adds value to their current workflows. The confidence issue could also be due to a lack of familiarity with the newly introduced visual representations. It is worthwhile to investigate this aspect in a longer-term field study.

With regard to detail-orientation (**EQ2**), we observed mixed behaviors. The participants had very different preferences regarding views, encodings, and detail-orientation. The slice-based views and the DVH plot seem to be the most helpful to the process based on all participants' feedback. It is "*a matter of habits*", according to P_2 . P_5 made use of the uncertainty distribution pop-up in the slice-based view, and the styling of the DVH lines guided her. Moreover, the uncertainty indicator seemed confusing to the participants as they struggled to "*grasp what was actually represented*" (P_2). P_4 thought that the RBE violin plot was difficult to read, although it could be useful to make decisions at a structure level. Most participants agreed that seeing the RBE uncertainty range is useful to their decision-making process. However, P_3 believes that using one model at a time for the calculation is a better approach. On the other hand, P_6 was skeptical about "*accept[ing] RBE calculations clinically*" overall. P_6 was the only participant to believe that the new views in the dashboard (beyond the conventional ones) would not be helpful to the decision-making process. The other participants agreed that, given complete RBE calculation results, the dashboard could provide new insights into the sense-making and decision-making processes.

Finally, we observed the users' decision-making processes during the study. Seemingly, the different views of the system were well integrated as users navigated the data at different detail levels. As anticipated, for an initial analysis, the participants were over-reliant on conventional views, shown in Figure 3.3 and Figure 3.7. Nonetheless, they employed the intrusive guidance provided in these views, shown in Figure 3.5 and explained in Figure 3.7. They regularly inspected the violin plot in Figure 3.6, which provided them with a comparative view of RBE uncertainty at a structure level, and the slice plot in Figure 3.4, which highlighted areas of difference at a slice and pixel level. On the other hand, the uncertainty indicator view (Figure 3.8), which provides quick guidance of interesting uncertainty deviations in slices, was not used efficiently. The users expressed confusion over how the plot derives its values, which highlighted the potential need for guidance explainability.

3.5.3 Use Case Example

In this section, we present a use case example. It has been conducted by one of our evaluation participants (*P5*, medical doctor) as part of the VDAR scenario. Further cases were conducted with the same patient data, and different participants analyzed their cases following different analytical processes and using different guidance levels and detail-orientations. The case below is just an example; no clinical inferences should be derived from it.

Participant *P5* chooses the two plans to compare with the dashboard. The brain stem is the most important structure to be spared. She hovers over its contour voxels to explore the dose value it receives. The uncertainty distribution comparison pops up. The plot provides per-voxel insight into the difference in robustness of the two plans.

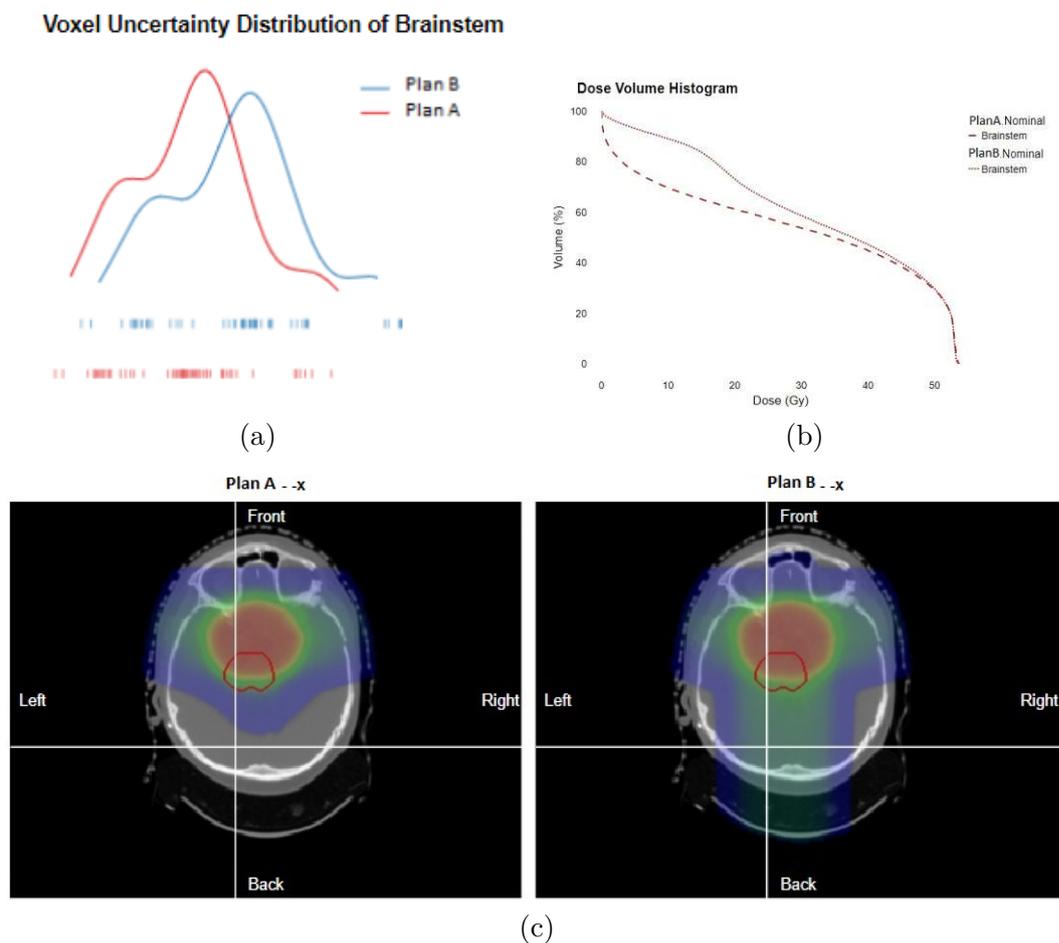


Figure 3.13: Comparison of two plans as conducted in a use case by *P5*, using the functionality of our dashboard. The contour colors follow the conventions in the clinical environment.

Plan B (Figure 3.13 (a) in blue) shows larger uncertainty. Subsequently, she looks at the comparison plot and sees an area of interest (per-structure analysis). She zooms into the brain stem region for a closer look and sets a customized clinical goal for the maximally accepted dose (54 Gy). She notices, in Figure 3.13 (b), the DVH line thickening, and she understands that the structure has received a higher dose than it should. She further inspects the dose distribution for both plans (per-slice analysis). She zooms in to the highest dose to see which plan is “*further out beyond 54 Gy*” and explores different set-up scenarios. She has indications that plan A could be more robust and may incur fewer risks to the brain stem, as seen in the slice analysis of Figure 3.13 (c). Plan A might be recommended in this case over plan B. The case example indicates that participant *P5* has employed all levels of detail (per voxel, structure, and slice) but not all views (i.e., the RBE violin plot and the uncertainty indicator were not used). Higher intrusiveness guidance was mostly preferred in her sense-making and decision-making process (and later positively assessed, as seen in Table 3.5).

3.6 Discussion

In this section, we outline the lessons learned from the feedback received in the user evaluation and during the uncertainty guidance dashboard design. We designed a visualization dashboard supported with uncertainty guidance mechanisms to improve the users’ confidence in their sense-making and decision-making process in PT planning, as expressed in section 3.3. The dashboard supports two tasks: the comparison and assessment of multiple PT plans (**T1**) and the analysis of their respective uncertainties, along with their impact on the treatment outcomes (**T2**). It is the first attempt to integrate different detail levels of guidance and degrees of intrusiveness into a clinical sense-making and decision-making process. We propose an evaluation framework to successfully measure confidence in guidance techniques by adapting two evaluation schemes to the needs of our application (ICE-T and SUS).

This work highlights the literature gap in developing usable visualization tools that provide guidance to users of PT planning (and other clinical) systems. The design process indicates that guidance can be a suitable strategy for supporting PT plan comparison. It *may generate more insights and add value to the current workflows (RQ1)*. It is particularly the case for more experienced users, such as *P3* and *P6*, who seem to prefer increased levels of guidance, i.e., highly intrusive guidance. In Table 3.5 *P3* and *P6* report increasing average points of 1.16 and 0.33, respectively, when using the full-guidance dashboard in comparison to using it without guidance. Regarding detail-orientated guidance, the prior analytical preferences of the users seem to matter.

The user evaluation also shows that *obtaining more insights does not necessarily mean increased user confidence (RQ2)*. Although *P3* reports an increase of 4 points in insight obtained, it corresponded to a reported 1.25 points decrease in confidence. The reduced confidence could be attributed to familiarity with the conventional decision-making process. At the same time, all detail-orientation levels seem to be needed, even though

different users might have different preferences regarding the views they want to use. Observing the sessions, we noticed that the participants relied more on conventional tools, such as those shown in Figure 3.3 and Figure 3.7, in exploring the therapy plans.

Participants in the user evaluation expressed interest in the comparative violin plots shown in Figure 3.6. The view enabled them to compare an essential dimension of uncertainties in plans, i.e., RBE values at a structure level, which was not visible to them before our work. This feature seems to be appreciated highly in PT decision making. *P3* reported that the tool is “*useful to get an overview of how the RBE uncertainties for critical organs compare.*” However, guidance explainability could boost users’ trust in the tool. To increase the value of the technique, an explainable guidance for conveying the underlying calculations to the users would be a helpful future addition and could increase confidence in the tool.

The variance in the participants’ experiences, as outlined in Table 3.2, contributed to the high variance in the evaluation results. Highly experienced domain experts might be more skeptical about the underlying uncertainty calculations or about adopting new approaches in their workflows. We recommend involving participants with varying experience levels throughout all design phases. From the elaborate feedback at the end of the sessions, we receive encouraging indications that our approach has value for PT planning. Additional studies with complete PT plan data are required to develop a more mature approach that could be adopted in clinical practice. *Onboarding techniques could be helpful in this case, as well as long-term studies (RQ3).*

The main limitation in developing and evaluating the dashboard was the restriction to use a synthetic dataset that cannot accurately account for more uncertainty components. Additional LET data could provide a more realistic evaluation setup, as the currently limited LET data affects the RBE uncertainty calculations and the final PT plan decision. Domain conventions and expectations contributed to a skeptical attitude toward the introduced dashboard. A further developed prototype or longer training could contribute to addressing this issue. It could highlight the visualization techniques’ value as complementary to conventional processes rather than being antagonistic. In addition, some of the representations in our dashboard have been designed to strictly follow conventions in the domain. However, the design could be improved and evaluated in the future according to best practices in the visualization domain, for example, colormaps for the dose overlay or the contours and encodings in the DVH plots. Also, analyzing multiple datasets, i.e., from different patients or tumor localizations, would be valuable.

In the design study, we relied on a mix of quantitative and qualitative approaches. The open comments are critical in informing the analysis of the results. In this study, each participant responded to six open questions and was encouraged to elaborate further on their responses. Although the open comments were constructive in informing our analysis, soliciting more dynamic open comments can further strengthen similar studies. For example, encouraging the participants to externalize their thought process can add value to the study. However, requesting continuous feedback can also come with a trade-off, as interrupting the normal analytical process of the participant might impact

their thinking and behavior. Furthermore, the self-reporting results of SUS might not present an accurate representation of indicators such as “confidence”. Thus, more complex evaluation methods are needed, as shown in later chapters of this dissertation.

Moreover, our observation is that guidance approaches fit very well with provenance approaches [SLSG16], which can provide additional knowledge on the path that leads to a decision. The proposed two-dimensional guidance concept can be generalized to other domains. We reckon that the approach applies to three areas in particular: weather forecasting, financial prediction, and material science. The data in these three domains embed multi-level uncertainties, which need to be explored at different detail levels, similar to PT.

Finally, as we clarified in subsection 3.2.1, our target users in this study are domain experts in PT. They have an existing trust relationship with the state-of-the-art system and models. They are also aware of the uncertainties introduced by the models. Their heuristics can introduce specific cognitive biases that impact their behavior. Our proposed two-dimensional mechanism could be generalized to other domains in contexts where the users are domain experts who similarly use a familiar state-of-the-art system and are aware of the uncertainties in the underlying models. This, however, would require further investigation.

Summary We presented a two-dimensional uncertainty guidance to support comparison and uncertainty analysis in PT planning visualization. We exemplify uncertainty guidance on a PT planning dashboard for plan comparison. Moreover, an evaluation measured confidence and insights from the use of uncertainty guidance. The evaluation yielded encouraging initial results and provided us with insights for forthcoming directions.

The chapter highlights several research opportunities. There is an opportunity to develop an evaluation framework to account for the domain-expertise factor. We think that it would be valuable to continue developing our proposed guidance taxonomy and mechanisms to reflect better the lessons learned from the current work. Additionally, it would be helpful to explore the potential of provenance techniques in informing uncertainty guidance. Another valuable future direction is the impact of explainable guidance on the user’s trust and confidence—especially in complex uncertainty scenarios. We anticipate that our approach (with modifications) can be generalized to other uncertainty visualization applications outside the medical domain, such as weather forecasting, financial prediction, or material sciences.

ConAn: Measuring and Evaluating User Confidence in Visual Data Analysis Under Uncertainty

This chapter addresses the need for metrics to objectively assess user confidence in the visual analysis process, and is based on the publication:

Maath Musleh, Davide Ceneda, Henry Ehlers, Renata G. Raidou, “ConAn: Measuring and Evaluating User Confidence in Visual Data Analysis Under Uncertainty”, in *Computer Graphics Forum*, Volume 44, No.1, 2025, e15272, doi:10.1111/cgf.15272.

User confidence plays an important role in guided visual data analysis scenarios, especially when uncertainties impact the analytical process. However, measuring confidence in practical scenarios remains an open challenge, as previous work relies primarily on self-reporting methods. In this work, we propose an approach to externalize and measure user confidence in visual analytics. The results suggest that our metrics improve the evaluation of user confidence compared to available approaches. In particular, we found a correlation between self-reported confidence and some of the proposed provenance network metrics. The quantitative results, though, do not show a statistically significant impact of the guidance on user confidence. An additional descriptive analysis suggests that guidance could impact users’ confidence and that the qualitative analysis of the provenance network topology can provide a comprehensive view of changes in user confidence. Our results indicate that our proposed metrics and the provenance network graph representation support the evaluation of user confidence and, subsequently, the effective development of guidance in VA.

4.1 Introduction

The assessment of *user confidence*, i.e. the “(self-)judgement of accuracy [...] after the completion of a task” [SKJ15, p.158], remains an underinvestigated topic [HQC⁺19], mostly because quantifying its qualitative traits is complex [Cof07]. Current research relies predominantly on self-reporting surveys to assess a technique’s impact on *user confidence* [DPBn07, HQC⁺19]. This might be particularly challenging in scenarios affected by uncertainties, which tend to have an impact on users’ confidence during a visual analytics (VA) session.

Uncertainties typically worsen *knowledge gaps*, i.e., the quantifiable difference between knowledge required to complete an analysis task and knowledge possessed by the user [CGM⁺17b, MMT⁺23]. Consequently, perceived *knowledge gaps* and uncertainties typically increase users’ doubts about their ability to complete a task successfully. As a countermeasure, VA researchers propose guidance mechanisms to mend knowledge gaps and improve users’ confidence during the sense-making and decision-making process. However, measuring the impact of uncertainty and, conversely, of guidance on users’ confidence remains an open challenge [HQC⁺19].

In VA, a way to analyze and understand the evolution of the underlying processes is by looking into provenance information. Past research proposes several metrics using provenance information to evaluate user trust and, simultaneously, to investigate the effects of guidance, such as the stability of user exploration [BLB21]. These proposed metrics remain insufficient to quantify and examine *user confidence*. Moreover, in previous work, there is no clear distinction between confidence and trust, and these terms are usually used interchangeably [SSK⁺16]. Such ambiguity complicates the assessment of *user confidence* in practical scenarios and hinders the development of solutions that help deal with uncertainty—among which is guidance.

Understanding the interlinking of guidance and user confidence allows us to tackle VA problems in several fields where multiple data uncertainties complicate the analytical process. For instance, in clinical decision-making scenarios or weather forecasting, decisions are impacted by several levels of uncertainty [RBS⁺18, MMT⁺23]. In such scenarios, decisions cannot be easily measured for accuracy a priori, i.e., the outcomes of a decision cannot be easily predicted. In these cases, we anticipate that guidance mechanisms can reduce perceived uncertainties and boost users’ confidence in their decisions.

Given that *confidence is a property of an attitude*, documenting provenance in the analytical process could provide a comprehensive framework of metrics that quantify user confidence under uncertainty. Specifically, one should consider a representation of the interaction provenance. We anticipate that representing provenance with a network graph allows us to harvest existing metrics for the objective quantification of user confidence while also constructing a topology that maximizes qualitative insight. In this work, we investigate the impact of guidance on user confidence within uncertain scenarios. We contribute:

- a conceptualisation of *user confidence* in VA — as opposed to *trust*;
- a model and metrics for the representation and probing of *interaction provenance* to analyze confidence in users’ sense-making and decision-making processes;
- a user study to validate the proposed model and metrics and to examine guidance impact on *user confidence*.

4.2 Related Work

Here, we review literature related to our work, namely the evaluation of user confidence and the use of analytic provenance in VA.

User Confidence Evaluation In the last decades, several evaluation methods have been designed focusing primarily on assessing users’ performance, user experience, such as usability, and VA algorithms [LBI⁺12]. Some approaches explicitly consider the impact of uncertainty and assess its impact on user confidence, mainly using self-assessment reports [AZB⁺15].

Hullman et al.’s [HQC⁺19] review of evaluation methods across 86 studies shows that most works rely on performance and satisfaction-based metrics to make a system’s appraisal, which is typically not sufficient to provide a detailed assessment of user confidence [SHK16]. Their work reveals an evident gap in value-driven evaluation methods owing to the fact that only a few approaches explicitly consider users’ *confidence*, whereas methods that do aim to study confidence predominantly rely on self-reporting, such as Likert scales or percentage estimation methods [ZAW⁺18, NWHL20, PRJ⁺23]. To fill this gap, Hullman et al. [HQC⁺19] recommend gathering information on the user’s reasoning towards a decision, as this information can be valuable in understanding how and why decisions taken during the analysis impact user confidence.

Wall et al. [WAM⁺19] propose a self-reporting qualitative survey based on Stasko’s [Sta14] value-driven method. Their multi-level heuristic-based approach called *ICE-T* provides an efficient tool to examine the utility of visualization, also considering, among other metrics, the effects of visualization on user confidence. Nonetheless, the proposed “confidence” heuristics only reflect user trust in the system. As we have seen, the interchangeability of trust and confidence is common in the evaluation of uncertainty visualisation [KSD⁺22, PVPM22]. A clear separation between the values of *user confidence* and *user trust* is crucial in studying visualizations and VA. Otherwise, study results could be deceptive.

Guo et al. [GDM⁺19] employ a user study to examine the impact of their novel uncertainty visualization technique on *user confidence*. While controlling for visual biases, they use a baseline to study the impact of confidence in decision making impacted by communicating uncertainty. Although efficient, their examination of user confidence was inferred rather than directly measured. Explicit metrics bolster an adequate assessment of the technique’s impact on *user confidence*. Moreover, they use partial scenarios for the study. Dimara

et al. [DBD18] emphasize the importance of real-life tasks and indirect self-reporting to study the impact on user confidence.

Guidance in visual analytics is defined as “a computer-assisted process that aims to actively resolve a knowledge gap encountered by users during an interactive visual analytics session” [CGM⁺17b]. Knowledge gap is the difference between the expected knowledge to complete a task and the user’s knowledge base. Considering guidance in VA, Ceneda et al. [CGM19b] study the impact of the guidance on user confidence using a self-reporting Likert-scale survey. Similar to Guo et al.’s [GDM⁺19] approach, they use a basic scenario to conduct their study. In addition, they use analytic provenance to correlate self-reported confidence to accuracy, as is common in the field. Thus, concerning *user confidence*, the effectiveness of the study approach diminishes where accuracy cannot be measured. To study system performance, they introduced several metrics. When slightly adjusted, their *distance* metric could measure *user confidence*. Studying analytical strategy and action quality is essential to infer user confidence.

In summary, existing literature largely confuses both concepts *user confidence* and *user trust* by using them interchangeably. Moreover, our review shows a stark lack of objective metrics to measure and study *user confidence*. The means used to study confidence are limited, and researchers elicit measures predominantly through *standard survey inputs*, such as self-assessment. In this work, we aim to surpass the traditional self-assessment methodology to evaluate users’ confidence in VA by repurposing a set of metrics used in social network analysis. However, as suggested by Hullman et al. [HQC⁺19], we focus on the analysis of interaction provenance as it can provide qualitative and quantitative insights into the user’s decision-making process.

Analytic Provenance Analytic provenance is a historical record of user reasoning with a VA tool [WAF16]. It is further classified into several categories based on the type of data recorded [RESC16]. Hullman et al. [HQC⁺19] distinguish between directly elicited, such as interaction logs, and derived measures, such as the number of interactions. For either, decision metrics in uncertainty visualization evaluation are scant. This exposes a gap in employing analytic provenance to assess uncertainty visualisation [JED⁺20, MSH⁺23]. The network of users’ low-level decisions or actions throughout the analytical process can potentially measure confidence. This approach can be connected to the use of implicit measures of interaction logs.

Indirect elicitation of eye-tracking and facial expression data provides valuable insight into users’ cognitive behavior [BKR⁺17, BVV⁺19, CAGM22]. Data on users’ physiological and emotional states can also be acquired using other biological methods and technologies such as electroencephalography (EEG) [HHH⁺11]. Nonetheless, these approaches demand substantial technological and logistical resources.

Boukhelifa et al. [BLB21] devise a tool that exploits provenance to measure user trust in guidance. Their approach provides an interesting path for examining *user confidence*. Their method records the user exploration path in XML log files to generate the metrics.

They analyzed the data generated by defined tasks rather than open exploration. This step adheres to the best practice suggested in the literature [DBD18, HQC⁺19]. They observe four indirect metrics to infer conclusions about user trust. At least one of the metrics is also helpful in measuring *user confidence*. Namely, observing exploration loops is closely related to the sense of self-confidence. Boukhelifa et al. [BLB21] needed to compare their observations to participants' self-reporting for conformity. However, they did not have access to such data; thus, they inferred it from other responses.

He et al. [HMH⁺21] also compare the metrics elicited from provenance to self-reported feedback. In their method, they record user interaction more rigorously, recording the screen and voice and asking participants to record verbal notes during the analysis. Nonetheless, this approach could disrupt the analysis process and, subsequently, bias the study. Thus, Nguyen et al. [NXB⁺16] suggest that collecting provenance should be *non-intrusive*. Several works present frameworks to collect and model provenance data to replicate the analysis process accurately [NWC⁺21, BFCA23]. To support their proposed framework, Nobre et al. [NWC⁺21] developed an open-source system to allow a non-intrusive method for studying users' analysis strategies. Nevertheless, the system does not consider the length of the event, which is a crucial factor in studying properties of user attitude, such as confidence.

Research demonstrates the usefulness of analytic provenance in explaining users' sense making [XAJK⁺15]. Visualizing recorded provenance as a network graph provides insight into users' cognitive process [OAB⁺17]. However, most such works rely on sequence pathways or hierarchical tree graphs [SLSG16, AKvM⁺18, CGL20, GMOB22]. The two most common provenance networks are *concept* and *history graphs* [XOW⁺20]. Huynh et al. [HEF⁺18] examine provenance data using network metrics. However, they also follow the trend of dedicating system artifacts as nodes in provenance networks.

Provenance is examined to evaluate systems [GL12, BGM19, NCWE22] with promises for developing provenance metrics [SHK16, FPH19, WNC⁺22]. Furthermore, research has relied on node-link hierarchical or event sequence graphs to represent provenance [BWD⁺19, DRM19, NTA⁺19]. The literature lacks work that applies conventional network graph metrics on provenance to semi-automate the evaluation of *user confidence* [BH19].

A comprehensive analysis of the impact of guidance on *user confidence* requires a multi-modal evaluation process. Both quantitative and qualitative approaches can provide complementary insights. Also, the analysis of the analytic provenance from a global and local perspective enables more comprehensive insights into the process. Metrics are needed to objectively quantify confidence, complementing the typical self-reporting measures. These metrics could, in turn, be complemented with a graph representation that allows a qualitative analysis of confidence evolution.

4.3 User Confidence vs. User Trust

As discussed in section 2.8, confidence and trust are terms often used interchangeably in visualization and VA. However, the logical semantics of the two concepts are different, as they entail different psychological and behavioral components. Understanding the distinction allows the design of VA solutions tailored to address specific needs related to trust and/or confidence. Moreover, clarifying and providing a better understanding of the semantics facilitates the formulation of explicit goals when designing and, subsequently, evaluating corresponding VA systems. Furthermore, the distinction between the two concepts can aid the VA designer in adequately pinpointing the exact reasons for system failures or pitfalls. We anticipate that distinguishing confidence from trust may ultimately support the development of effective evaluation metrics to measure confidence and trust in VA.

Sociology distinguishes *confidence* and *trust* by their relationship to *agency*, the presence of *risk*, and the *consequence* [MW09]. Confidence is an act of giving out agency, according to Luhman [Gam88, Luh05]. On the one hand, in *confidence*, the user attributes the error in the decision externally, for example, incomplete or missing data. On the other hand, in *trust*, the blame is internalized, i.e., it relates to one’s analytical reasoning. For example, a driver trusts other drivers on the road based on preconceived stereotypes. The driver accepts the risk as they do not have *evidence* of each person’s driving skills. Thus, if the assessment of other drivers proved wrong, the blame is internalized on the preconceived stereotypes. While driving, however, the driver analyzes traffic signs to make decisions: the green traffic light means you can drive and continue driving with confidence. If there is an issue with the traffic light, displaying green instead of red, causing an accident, the driver will blame the traffic light. Thus, the blame is external. We summarise the main characteristics that differentiate the two concepts of confidence and trust based on factors from sociology literature in Table 4.1.

User confidence is a *self-certainty* which relates to the subjective judgment of the accuracy of one’s decision based on specific acquired knowledge and a frame of reference [MKLS13, SKJ15]. It describes a relationship between oneself and is intolerant to risk. This means that when the sense of risk increases, confidence decreases. For instance, when a user faces situations characterized by high risk, for example, high uncertainty caused by incomplete

Table 4.1: Outlining the differences between *confidence* and *trust* based on four factors: risk, agency, evidence, and consequence.

Confidence	Trust
Intolerant to <i>uncertainties</i> and <i>risk</i>	Tolerates <i>uncertainties</i> and <i>risk</i>
User gives <i>agency</i> to system	User retains <i>agency</i>
Requires strong <i>evidence</i> to be boosted	Can be boosted with no or little <i>evidence</i>
User blames <i>consequent error</i> on the system	User blames <i>consequent error</i> on misplaced trust/judgement

or ambiguous information, they may experience a decline in confidence in their ability to make accurate judgments—especially when considering the potential consequences of said decisions. Although users seek internal assurance about the accuracy of their decision, the decision’s accuracy does not necessarily correlate with the self-accuracy of judgement [SHK16], i.e., one might be over-confident about their judgment.

User trust refers to the property of the user’s attitude *towards an external party*, i.e., the visualization system and data, and can be seen as the subjective judgment of the accuracy of the system’s information, decisions, or options without (or with minimal) evidence, for example, a calculation method. Contrary to confidence, trust is highly tolerable to risk. This means that the user may believe in the correctness of a representation without necessarily seeking evidence. A user’s trust can tolerate certain levels of risk when there is a perceived benefit or value in trusting the system. For instance, when the system fails to support the analytical process, the user may rather blame their own judgment of the system’s competency—they believe their trust has been misplaced.

Within a system, *confidence* and *trust* interact and impact each other dynamically, especially in the presence of risk. In such cases, a lack of confidence in their own decision-making processes may lead users to rely heavily on trusted systems. Conversely, if users trust a system but encounter unexpected errors, their confidence may be impacted, leading to reevaluating their trust in the system. For example, a student is deciding on their specialization in a university and analyzes several data sources and uncertainties in the process. They obtain data on best-paying jobs, which they *trust* to be accurate without investigating this further. After analyzing the available data, the person decides on the optimal specialization with a certain degree of *confidence*. This confidence is impacted by the amount of data and the person’s ability to analyze the available data and reduce the perceived uncertainties.

User uncertainty presents another level of uncertainty directly connected to confidence and trust. Within a guidance-supported VA system, uncertainty in decision making reflects a lack of confidence (the user is not certain about their actions), while uncertainty in guidance pertains to a lack of trust (the user is not certain about the system). These attitudes manifest differently. Distrust in guidance may lead users to disregard the latter, but a lack of confidence may manifest as indecisiveness or frequent changes in decisions. Both manifestations can be tracked and documented within an analytical process to support further understanding of the underlying processes of user confidence with regard to guidance in an uncertain analytical scenario.

4.4 Hypothesis and Research Questions

We hypothesize that *provenance network graphs enable an objective measuring of user confidence and a better understanding of how confidence is impacted by guidance and uncertainties*. Our work investigates the following *research questions*:

RQ1 How can we *measure user confidence* in VA under uncertainty?

RQ2 How can we exploit *analytics provenance* to better understand *user confidence* throughout the analysis process?

RQ3 How does guidance impact *user confidence* throughout the analytical process?

In this work, we propose metrics to measure user confidence. We also examine the impact of guidance on user confidence in VA under uncertainty. We do so by repurposing social network analysis (SNA) metrics that measure analytics provenance, converging qualitative and quantitative methodologies. Analytical tasks often involve multiple steps, iterations, and interactions among various components such as data sources, visualization tools, algorithms, and user actions. Interaction provenance refers to the history or lineage of interactions within an analytical process, detailing how users have interacted with data, tools, and visualizations. A network graph can visually depict these relationships, providing insights into the flow of information, dependencies between actions, and the overall structure of the analytical workflow. Thus, representing provenance as a network involves depicting these interactions as nodes (points) and their relationships as edges (connections). Each node may represent a specific action or event in the analytical process, while arcs denote the flow or sequence of interactions. Using a network representation, one can capture the complexity and interdependencies inherent in the analytical process. For instance, as users' confidence in their decision depends on their perception of the sense-making process [SHK16], a provenance network's local and global topology can provide new insights into the user's underlying thought and analytical processes.

To measure the adequacy of our approach, we designed and developed a user study with which to explore the effects of guidance to possibly reduce uncertainty and boost *user confidence* based on a VA scenario supported by a dashboard. We propose provenance metrics and probe users to self-report confidence in two surveys to address **RQ1**. We devise survey questions to measure subjective confidence distinctly from trust, and we confront the results obtained with our metrics against the user's self-assessed confidence. In this regard, we administer an intermedial survey and an end-of-session survey. Furthermore, in representing provenance as networks and exploiting metrics from the domain of SNA, we develop a framework to understand better *user confidence*. This aims to answer **RQ2**. Finally, we show that network graph metrics and representation provide a better understanding of what factors impact *user confidence*. Analyzing confidence values propagated by the provenance network against the self-reporting results helps us investigate **RQ3**.

4.5 Dashboard for Farming as a Proof-of-Concept Scenario

To showcase the use of the SNA metrics, we created a dashboard for eco-farming data analysis. We chose this scenario because eco-farming entails many different sources of uncertainty, making decisions complex for farmers. In such scenarios, the farmer undergoes several shifts in trust and confidence based on several unexpected factors, such as weather

conditions, prices, and so on. Considering uncertainties in the ecological system in farming decision making resembles the accounting for human biological system uncertainties and shifts in clinical decision making. There is a full spectrum of information that needs to be considered, and confidence in navigating this information space is necessary, i.e., one should be confident of one's capabilities in making the right decision. This scenario was chosen due to the presence of multiple uncertainties and its suitability for a potential study with the general population without requiring expert knowledge of farming. It allows us to investigate our scenario with a larger audience, enabling us to measure confidence quantitatively.

Farmers start planning for the season a few weeks in advance. The main goal of farmers is to decide about *which crops to plant* and *when to plant them*. In this sense, many different factors need to be considered. For instance, *seasonal temperatures* guide what crops they will plant. Typically, small eco-farms maintain a direct relationship with the consumer. Thus, *consumers' preferences* influence the decision-making process. Moreover, farmers must consider the best *crop variety*. Unlike chemical farming, eco-farming depends on the season's high *diversity of crops*. Planting a single crop (i.e., monoculture farming) is not environmentally sustainable. Farmers must consider the planting cycle within a planting bed to reduce the risk of diseases. Furthermore, farmers alternate (each year) within a specific bed, i.e., a section of land, between five types of crops: fruits (including flowers and bulbs), greens (leaves), roots, legumes (including seeds), and tubers. In small farms, a farmer must make several decisions based on experience and expert consultation before deciding what *crop combination* to plant to reduce disease risk. This information is not readily accessible. Profit is essential for small farms to remain sustainable. Thus, *price fluctuations* and market supply play an important role in deciding *when to start planting*. An *early planting*, within the acceptable window, could result in more profits despite a higher risk.

4.5.1 Available Agricultural and Weather Data

We collected agricultural and weather data from several sources. We use data from the Palestinian mountain region.

Crops Data: We obtained data including crops' best and worst companions, planting periods, and crop information from two farming booklets [Kor12, Om 18]. The booklets provide qualitative information on which crops benefit each other and how crops benefit each other. The planting period includes ranges of maximum and minimum temperatures (in °C) and ideal high and low temperatures (in °C) to plant, grow, and cultivate crops. Crop information includes data on the optimal distance (in cm) between seeds/seedlings and the weeks of cultivation.

Historical Weather Data: We obtained daily weather records from Visual Crossing [Cor23] for Jerusalem, Palestine. The data span 20 years from January 1, 2003, and include several weather features. For our scenario, we focus on four main features,

which are the maximum temperature, minimum temperature, average temperature (in °C), and precipitation (in mm).

Historical Prices Data: We obtained data on crop prices for each month of the year from the Palestinian Central Bureau of Statistics (PCBS) [Pal23]. We consulted local farmers to complete the missing data based on their experiences.

4.5.2 Study Tasks

We anticipate that the upcoming user study participants are students with a general interest but minimal or no experience in farming. The participants are tasked to decide on which crops to plant in their hypothetical farm over the course of two sessions (one guided and one unguided). As seen in Figure 4.1, each session has four equal sub-tasks. For each sub-task, users make three sub-decisions on the *main crop* (SD1), the *companion crop* (SD2), and the *planting week* (SD3). To make their decision, they engage with the interactive dashboard in a sense-making loop.

We outline two task examples for the user evaluation sessions. The tasks are formulated as questions demanding answers through user analytical and exploratory activity. Both tasks have similar levels of difficulty, with different expected decisions.

Task 1: You are in year 4, around April. You have land that contains four planting beds, each $16m^2$ in size. In previous years, you planted the following in the beds: in bed 1, you planted legumes, roots, and greens; In bed 2, you planted roots, greens, and tubers; in bed 3, you planted fruits, tubers, and legumes; and in bed 4, you are planting for the first time. You want to plant the season and sell 60% of the crops in order to maximize profits. What combination of crops will you plant in each bed? And in which week will you start planting each bed?

Task 2: You are in year 4, around October. You have land that contains four planting beds, each $10m^2$ in size. In previous years, you planted the following in the beds: in bed 1, you planted tubers, legumes, and roots; in bed 2, you planted fruits, tubers,

Task Decision			
Sub-task Decision			
	Sub-decision 1 (SD1)	Sub-decision 2 (SD2)	Sub-decision 3 (SD3)
Bed 1	Beet x ▾	Beet x ▾	Week 3 x ▾
Bed 2	Beet x ▾	Companion ▾	Plant on ▾
Bed 3	Main Crop ▾	Companion ▾	Plant on ▾
Bed 4	Main Crop ▾	Companion ▾	Plant on ▾

Figure 4.1: The panel where the user makes decisions in the dashboard outlining three levels of decision: sub-decisions, sub-task decisions, and task decisions.

and legumes; in bed 3, you are planting for the first time; and in bed 4, you planted legumes, roots, and greens. You want to plant the season and sell 60% of the yield and keep the rest for your household. What combination of crops will you plant in each bed? And in which week will you start planting each bed?

4.5.3 Eco-Farming Dashboard

A guidance-supported VA system can provide inexperienced local farmers with tools to make optimal farming decisions for the season and more sustainable farming practices in the long run. The dashboard provides the task list criteria for the month, area of land, and previous planting cycles. The user must complete several *sub-decisions (SDs)* in four *sub-tasks (STs)* to fulfill the task, as seen in the decision panel in Figure 4.1.

The dashboard idea is an upgrade of a previous ontology-supported dashboard, Agritology [GM21]. It is clustered into three views presented in Figure 4.2, Figure 4.3, and Figure 4.4. The *weather analysis view* (Figure 4.2) allows probing weather-related data (precipitation and minimum, maximum, average temperature) for a desired span of years in a timeline view on a weekly/monthly basis. Users can get a summary of the weather with respect to past data. The *crop information view* (Figure 4.3) allows probing at which time of the year

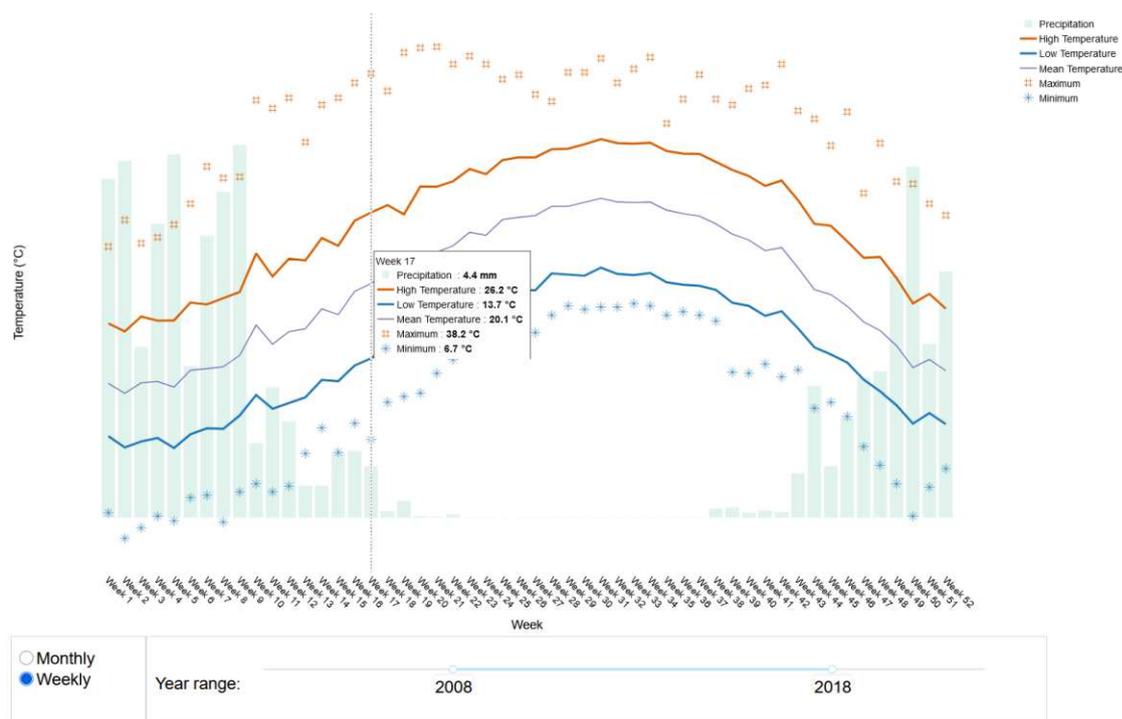


Figure 4.2: Weather Analysis View of our proof-of-concept dashboard for eco-farming: visualizes the temperatures and rainfall data for each month/week of the year.

4. CONAN: MEASURING AND EVALUATING USER CONFIDENCE IN VISUAL DATA ANALYSIS UNDER UNCERTAINTY

crops should be planted and harvested, using a heatmap. For each month (column), the respective crops (rows) are indicated. The view can be filtered to provide more specific information on suitable temperature ranges for crop planting, growing, and harvesting.

Additionally, a second view presents a horizontal bar plot that allows users to obtain information on the suitable distance between each seed. The view can be filtered to obtain information on the expected yield range for each crop and historical crop prices for each month for each crop. This information allows the user to calculate the expected profit range of their choices. The *crop companion view* (Figure 4.4) allows visualizing in a heatmap which crops (column) benefit which (row) in order to minimize risks. The heatmap can also be filtered to specific benefits: insect repellent, disease control, or crop yield. For example, mint benefits onion (green), while green beans are detrimental to onion (purple).

The aforementioned visual representations (timelines and heatmaps) are chosen for simplicity and affordability with large audiences. Additionally, we have another panel (shown in Figure 4.1) where users can record their decisions. Finally, we track users' actions by recording each action they take on the dashboard based on a previously outlined range of afforded actions. This is recorded in a dictionary of actions and timestamps. These actions are the third level of user activity in VA as described by Gotz and Zhou [GZ09].



Figure 4.3: Crop Information View of our proof-of-concept dashboard for eco-farming: visualizes the farming calendar of planting and cultivation, and the ideal distance between seeds or seedlings.

4.5. Dashboard for Farming as a Proof-of-Concept Scenario



Figure 4.4: Crop companion view of our proof-of-concept dashboard for eco-farming: visualizes the suitability of planting crops together.

Tracking user action on a dashboard could be disruptive. There are several techniques to record the user interaction at different levels of the cognitive process and disruptiveness [RESC16, XOW⁺20, dOBD⁺22]. In our dashboard, we organize views and eliminate certain elements from plots to drive the user to expose their exploration steps explicitly to avoid disruptiveness in recording the interaction provenance. For instance, we put views in clusters based on the sub-task they fulfill. We also remove the y-axis ticks to track user analysis as they hover over the plot. This way, users must click on a tab to reach a view cluster or a radio button to explore a specific view. This allows us to know which cluster and view is being explored.

Regarding guidance, as seen in Figure 4.5, the dashboard prescribes which crops or weeks the user should explore or analyze. This prescription is illustrated as a list on a side panel related to the user's current analytical view. The new suggestion is highlighted when the guidance list is updated based on the user's actions. The view tab is also highlighted when guidance becomes available for the other analytical views. This guidance can be switched off on demand.

4. CONAN: MEASURING AND EVALUATING USER CONFIDENCE IN VISUAL DATA ANALYSIS UNDER UNCERTAINTY

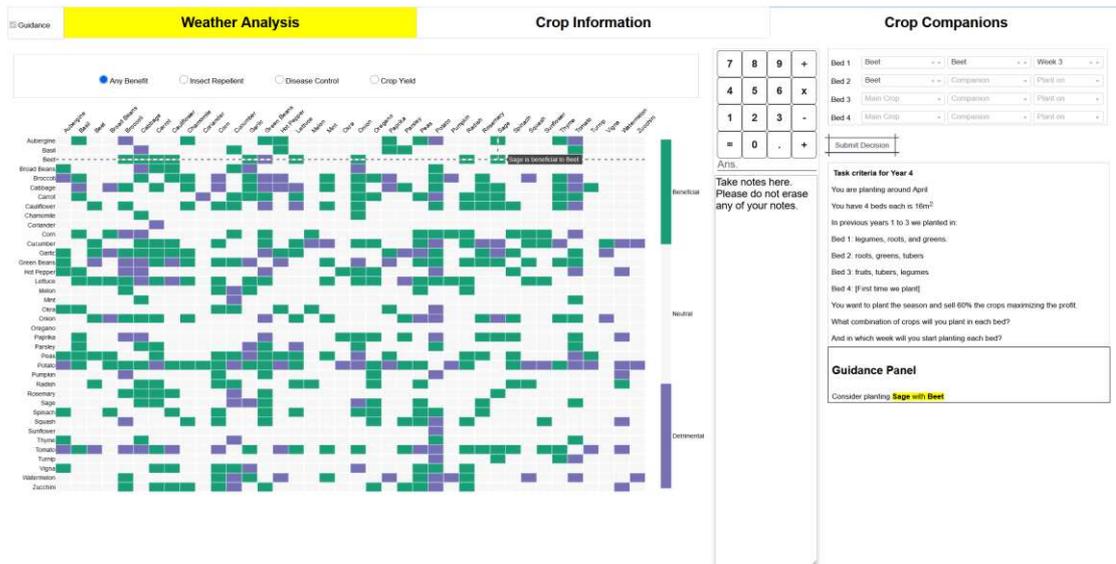


Figure 4.5: Entire view of our proof-of-concept dashboard for eco-farming: on the bottom-right box of the panel guidance is exemplified with a highlight reference to the Weather Analysis tab on the top left.

4.6 Provenance Network Graph

We represent the provenance data as a network graph and repurpose network metrics to garner insights into *user confidence*. A *network graph* is constructed of *nodes* that represent entities connected by *arcs* that represent a relationship between the nodes. When it is directed, a network edge is commonly known as an *arc*.

Our provenance network graph (ProvNet) illustrates the transmission of knowledge (*arcs*) from one user action (*nodes*) to another, representing the user’s sense-making and decision-making process. It is a fully connected network. The ProvNet is a visual, mathematical representation of a connected system of (analytical) user interactions with a VA dashboard. An example is shown in Figure 4.6.

Moreover, we draw the ProvNet in a circular layout as it offers a better understanding of the inter-cluster flow of knowledge. This layout allows us to detect patterns better in the analytical process. We set the clusters a priori based on views described in subsection 4.5.3. Each view supports the user in making one of the sub-decisions outlined in Figure 4.1. Subsequently, all actions triggered by the user in the analysis of a view cluster are classified within that cluster. We use a letter indicator for each cluster as outlined in Table 4.2 to simplify the labeling of the actions and avoid cluttering the network graph.

Nodes We use *Actions* as nodes to create the ProvNet, as outlined in Table 4.2. These actions are specific to our dashboard and should not be seen as an exhaustive list of

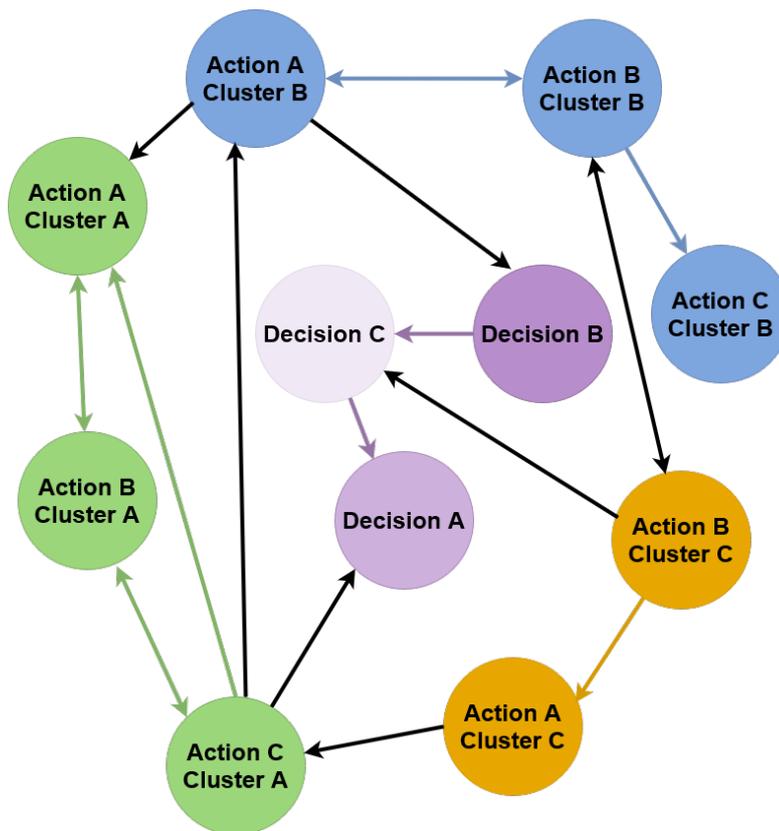


Figure 4.6: An illustration of a provenance network graph. The actions taken by the user are predefined into clusters that correspond to different levels of sense making and decision making (indicated by the colors). Each cluster of actions includes varied numbers and types of actions. They are all connected with directed arrows representing the flow of action and knowledge from one node to another. The opacity in the **Decision** nodes represents the percentage of decisions following the guidance provided.

all potential actions a user might take in any VA system. For a different scenario, the VA designer would need to define all the actions afforded by their system on the same granularity. In any case, an initial requirement to collect the interaction provenance data is to pre-define all possible *actions* and classify them into clusters based on the sub-tasks. In our study, the *actions* code mimics the following naming model:

[view cluster]_[action type][action type variation in a cluster]

Each node carries a *time* variable, representing the sum of time taken before triggering the following action. The opacity in the decision nodes encodes the percentage of decisions influenced by the guidance. For example, a user may not accept all guidance provided. Only 30% of their decisions might match the guidance provided. The percentage is calculated exponentially to allow enough variance in the opacity for observation. In the

4. CONAN: MEASURING AND EVALUATING USER CONFIDENCE IN VISUAL DATA ANALYSIS UNDER UNCERTAINTY

Table 4.2: Abbreviations used for the actions of the network. The colors indicate a distinct cluster of each action. The rows are the different types of actions, and the columns are the different view clusters. The letter abbreviations are combined to create the abbreviated name of each node. The numbers in (\times) denote the distinct variations of an action type. Each variation is represented with a separate node in the provenance network graph. Thus, the maximum number of nodes in any network is 24 nodes: five in cluster **A**, ten in cluster **B**, six in **C**, and three in the decision cluster **D**.

Clusters \ Actions	(A) Weather Analysis	(B) Crop Information	(C) Crop Companions
(V) View Exploration	a_v	b_v	c_v
(F) Filtering	a_f (x3)	b_f (x7)	c_f (x4)
(N) Note-taking	a_n	b_n	c_n
(C) Calculating	x	b_c	x
(D) Decision	a_d	b_d	c_d

example mentioned, we calculate it as $\frac{100}{10} * (10^{0.3})$ which gives us an opacity value of 20%. Although typically 12 sub-decisions are submitted (4 beds \times 3 decisions), a user may make more than 12 sub-decisions before they submit a final decision, for instance, a decision can be changed. This case is shown in Figure 4.6, where decision C is more transparent, i.e., less guidance was accepted during this decision.

Arcs The network’s ties (arcs), in our case, represent the knowledge flow between each action (node) during the analytical process. In VA, the sequence of actions is often complex, resembling a network of activities rather than a linear sequence of events. Moreover, the activity is not initiated arbitrarily but is influenced by the previous activity. In our context, the arcs represent the flow of the analytical process, implicitly tied to the knowledge gained by users from specific actions within the VA system. Each action undertaken in the VA process leads to acquiring new insights, shaping subsequent actions. This transmission of knowledge guides the user through the analytical process. Thus, arcs represent a functional linking [EK15] that describes the network’s knowledge flow. Moreover, each arc is characterized by the frequency of sequence repetition. We color arcs to denote inter-cluster (grey) and intra-cluster (cluster color) interactions, as seen in the arcs in Figure 4.6.

From User Log to ProvNet We process the provenance by reading the logged dictionary of actions and their timestamps. Each action is processed as a node. We deduce the clusters from the logged action name and sum up each unique action’s time length. For the guided sessions, we calculate the percentage of decisions following guidance coded in the logged provenance to adjust the opacity of the decision cluster **D** nodes. Then, we construct a node table with the timestamp as a unique index. The table contains data on *action node*, *action type*, *action cluster*, and *opacity*. Subsequently,

nodes are paired based on consequential timestamps. In essence, we use the timestamp and action node columns to construct a pairs table that builds the arcs of the network graph. We assign the arc weight as the occurrence count for each unique node pair. Figure 4.6 illustrates an example of such a network graph.

4.7 Quantifying Confidence

Although Social Network Analysis (SNA) emerged in social sciences, scholars in various sciences find its metrics helpful to study and observe different non-social networks [MBBM22]. We use metrics from SNA to observe the user flow of actions in VA under uncertainty to study and assess the variability of *user confidence*. We observe the global metrics of the network to infer user confidence propagated by the analysis process. We also observe the local metrics of the network to study user confidence at the node level. The latter enables us to investigate how guidance impacts and influences user confidence. The objective metrics are drawn from the network graph on three levels: global characteristics, global topology, and local topology. Below, we explain these metrics in detail.

4.7.1 Global Characteristic Metrics

We aim to measure *user's confidence* in their analytical decision-making process under uncertainty using global network characteristics (**RQ1**). Specifically, we compute two metrics, (linear) *network diameter* and (linear) *network size* [MBBM22]. A linear network denotes the network representation when the actions are split into the event instances; thus, it represents a linear graph of the user interaction. Each node, except for the *start* and the *end*, has one incoming and one outgoing arc.

(Linear) network diameter The typical understanding of *diameter* in network analysis is the shortest path between the two most distant nodes in the network [MBBM22]. With a directed weighted graph, the diameter will thus equal the total time spent from the start until the end of the analytical process. For this, we calculate the time spent at each action cluster by summing up the time of each action for each cluster. Longer times spent making a decision reflect high uncertainty in the decision, as *reaction* is, in fact, inversely correlated to confidence [SHK16]. However, this metric should be interpreted in connection with the number of actions. Moreover, a time threshold should be set a priori in time-sensitive analytical processes, such as in clinical decision making.

(Linear) network size We count the number of actions taken by the user. We assume that more actions in an analytical process could indicate that the user is uncertain and seeking more confidence before making a decision. However, as explained above, the number of actions should be interpreted with analytical time. A shorter overall time with more actions performed could indicate more confidence than struggling for longer with certain actions. On the other hand, a longer overall time with fewer actions could result from uncertainty in sense making.

4.7.2 Global Topology Metrics

We consider two graph metrics: *centralisation* [Fre79] and *density*, illustrated in Figure 4.7. Both metrics examine the cohesion and compactness of a network. These two metrics are the most popular metrics for studying the global properties of a network.

Centralization This metric presents an understanding of equality within the network. Freeman’s network centralisation [Fre79] helps examine small networks such as the interactions network in our study, measuring graph centrality and the extent of its cohesion around a central node. The centralization value is between 0, a complete graph, and 1, a star, the most centralized network. High network centrality tells us that fewer actions within the visual analysis are central to the user’s decision making, which could demonstrate more confidence in the decision-making process as it indicates a cohesive analytical process.

We assume that the different views (clusters) in an efficient visual analytics system are equally important to the sense-making process. Thus, in our case, we assume that the accepted level of centralization to infer the typical sense-making process is around 0.333. We derive this value by calculating $\frac{1}{x_s}$, where x_s denotes the number of sense-making clusters. A lower value than the selected threshold could indicate a lack of confidence in the process, and a higher value indicates high confidence.

Graph Density Low network density indicates a free-flow analytical process. If the network density (N_d) equals $\frac{1}{n}$, where n is the number of nodes in the network, then the analytical process is perfectly linear [EK15]. Thus, $\frac{1}{n} < N_d \leq 1$.

Nevertheless, the *density* metric cannot be interpreted individually. One must also co-examine the *network diameter* (time) variable and network centralization. A low density with a lengthy exploration time indicates that the user needs support, as it could potentially indicate low confidence at the sub-task level. Conversely, low density with a short exploration time could indicate high confidence. High density with a lengthy exploration time may result from a struggle to understand the main task, indicating a lack of confidence. High density with a short exploration time is the least likely scenario; however, it could result from extreme familiarity with a process.

The total number of possible connections in a graph is $G_T = \frac{N(N-1)}{2}$; where N is the total number of nodes. In a *wheel graph*, a central node is connected to all nodes connected in a circular structure [EK15]. We deem the wheel structure for each cluster to resemble a confident analytical process, given that all clusters are connected. Thus, the total connections in such a structure are $G_w = (x - 1) + \sum_{c=1}^x 2(N_c - 1)$; where x is the total number of clusters, and N_c is the number of nodes in a given cluster. In our case, with four clusters containing [5, 10, 6, 3] nodes, we expect the density for a confident analytical process to be $\frac{G_w(43)}{G_T(276)} = 0.156$. We deem the maximum number of acceptable densities

to be when the clusters are fully connected. Thus, $D_{max} = \frac{t(x-1) + \sum_{c=1}^x 2(N_c-1)}{G_T}$; where t denotes the number of sub-tasks, i.e., four in our case study. Thus, in our case, the

accepted D_{max} is 0.188. We accept a star structure (one central node connected to all nodes [EK15]) of each cluster, given that all clusters are connected in each sub-task, as the minimum expected density of a confident analytical process. Thus, $D_{min} = \frac{t(x-1)+(N-x)}{G_T}$. In our case, the accepted D_{min} is 0.116.

4.7.3 Local Topology Metrics

The network can provide valuable insight into the analytical process [HEF⁺18]. It can also provide reasons for confidence variability at a local level while indicating which actions are important to the user or prominent in the analytical process. We use two centrality measures to extract local topology measures: *betweenness* and *in-degree* centralities [Gol15, Hev18, MBBM22] (see Figure 4.7). Additionally, we analyze the ProvNet to understand the reasons and structures of the confidence level.

Betweenness Centrality This metric indicates node importance. It is calculated as the total number of shortest paths that pass through a node divided by the total number of shortest paths for each pair of nodes in a network. The calculated value is then normalized on a scale of 0 to 1. If the node has a betweenness closer to 1, it indicates that the given action boosts the user’s confidence. The node with the highest betweenness value controls the flow of knowledge in the analysis process [Fre79]. Hence, its removal could disrupt the analytical process. The metric ultimately points to the actions that are important to the sense making.

In-degree Centrality This measure indicates a node’s prominence. It is calculated for each node as the number of connections of incoming arcs. We can examine each guidance node’s prominence level to determine if it is central to the analytical process. Nodes with high in-degree centrality may provide confidence to the user when making a

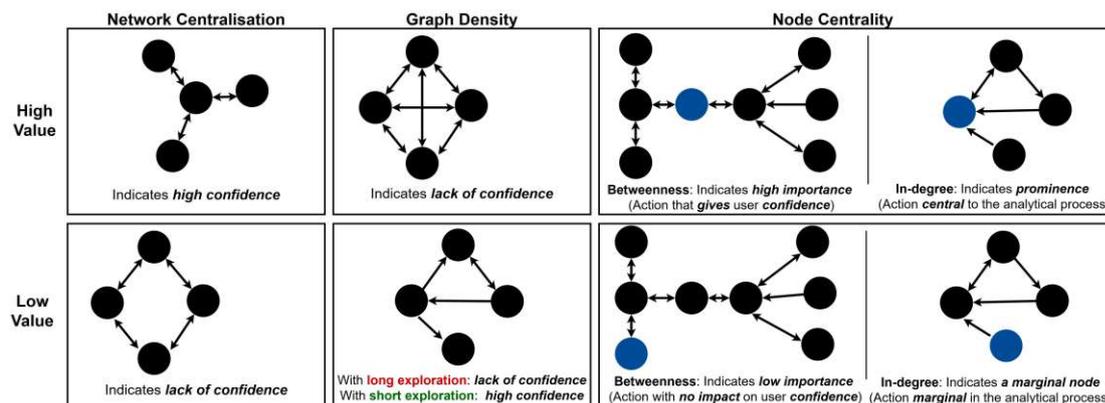


Figure 4.7: An abstract illustration of the global and local topology metrics that we propose to use in analyzing the ProvNet.

decision. However, a node might not be important to the sense-making process and has less impact on the process if it has a lower betweenness centrality value.

4.8 Evaluation Methodology

We structure the user study into three phases. In the first phase, the participant completes a tutorial. In the second phase, they use the dashboard to make a task decision and complete a survey. Finally, we asked participants to complete a second task and a final survey in the third phase. We collect data on participants' decisions in each session, their interaction with the dashboard, with timestamps for each action, and the notes they might have typed during the session. We conducted a pilot study with three colleagues to assess the tasks' difficulty and feasibility.

To measure confidence, we apply a mixed approach: 1) we integrate a classical self-reporting survey approach and combine it with 2) a provenance analysis to validate our proposed metrics and study the impact of guidance. Thus, we compare and correlate the objective to the subjective quantification of user confidence. We measure *user confidence*, quality of user experience, to direct our evaluation. Additionally, we aim to understand the impact of guidance, a technique applied in the analysis under uncertainty. We use self-reporting measures and both direct and derived implicit measures from the ProvNet graph. We present the self-reporting measures at the beginning and end of the session. The following hypotheses are tested in this study:

Hypothesis 1a (H1a): The total time and number of actions correlate negatively with self-reported confidence.

Hypothesis 1b (H1b): The impact of guidance is reflected in confidence by a significant effect on *network diameter* and *size*.

Hypothesis 2a (H2a): *Centralization* as a global topology metric can predict the users' self-reported confidence in decision making.

Hypothesis 2b (H2b): *Density* as a global topology metric can predict the users' self-reported confidence in decision making.

Hypothesis 2c (H2c): Guidance significantly impacts the *global topology metrics*.

4.8.1 Survey Design

The control survey (Survey 0) captures the user's subjective prediction of future performance and decision accuracy. We use it as an initial control point to measure whether the user's confidence improved. In contrast, the in-session pop-up questions track the changes in confidence levels throughout a session. We use the results of the first post-session survey (Survey 1) and second post-session survey (Survey 2) to keep track of the overall progress in confidence levels and the impact of the guidance on confidence variability. We also correlate the subjective metrics of the self-reported surveys to objective metrics from the network.

We draw from social psychology literature to construct the survey. In social psychology, confidence is typically assessed with a self-reported survey or judgment of task completion accuracy [SKJ15]. Therefore, we collect and analyze two types of data: self-reported confidence and user decisions. The user completes three primary 5-point Likert-scale surveys to self-assess their confidence: **Survey 0** before the start of the study on self-efficacy (9 questions), **Survey 1** following the first session (9 questions), and **Survey 2** following the second session (11 questions). Furthermore, we assess each participant's confidence in the decisions periodically during the sessions by completing one 5-point pop-up question (**Intermedial survey**) upon making a complete set of decisions for a bed, i.e., when a participant decides what pair of crops to plant in a bed. The number of self-assessments is not fixed a priori, and each participant will respond to a different number of pop-up questions depending on the number of times they change their decisions.

We standardized the language in probing the questions and the response scale for all three surveys and the in-session pop-up questions (Intermedial survey) based on best practices [SKJ15]. In the three primary surveys, five questions assess users' confidence in domain knowledge and tasks, and four assess their confidence in using the dashboard. This enables us to locate sources of confidence (or lack thereof) when analyzing the results. We ask two additional questions in the second post-session survey (**Survey 2**) to assess if the participants' confidence is affected by the guidance tool.

4.8.2 User Study Participants and Protocol

We invited 20 participants to conduct the study, and they were offered twenty euros for their participation. The participants were recruited by announcing the study in different master classes in the computer science department. Before the sessions, we sent a two-page reference on steps and considerations in the decision-making process in sustainable farming and asked them to get familiar with the domain. During the single sessions, we asked participants to sign a consent form and illustrated the study purpose and considerations for making decisions in sustainable farming. We emphasized that the study is testing participants' decision-making process rather than the usability of the dashboard. Afterward, participants followed a short tutorial on using the dashboard. We curated the tutorial as external onboarding [SCW⁺22].

We did not interfere in the study to avoid biasing participants' confidence. However, we were present in the room to tackle any technical issues that might arise with the use of the dashboard. We provided the participants with a printout of the reference document on the decision-making steps. We booked 1.5 hours for each session, but we did not limit participants to a timeline. Thus, in the case of consecutive sessions, we prepared a backup plan to conduct a session with similar conditions in case the first session would be over 1.5 hours. We recorded the screen during the study as a backup to the collected data and to offer potentially additional sources for a qualitative evaluation. Our participants were a mix of Bachelor's and Master's male and female Computer Science students. We did not collect the participants' demographics as these are not necessary for our analysis. We do so to protect the participants' anonymity and avoid study bias in accordance

with best practices [FGD⁺16] and the directives of the European Union General Data Protection Regulation and the Austrian Data Protection Authority.

4.9 Results

Each participant ($n=20$) completed two sessions in a randomized order, one guided and another unguided. Thus, we have a sample of 40 sessions. We removed one outlier that positively skewed the results (finally, $n=19$). Participant (42864671) spent 102.2 minutes analyzing the first task (unguided), compared to an average of 27.17 minutes in the other unguided sessions. After spending more than 100 minutes, she was reminded that she still had one more session to go. She explicitly stated that she misunderstood the instructions. Subsequently, she wrapped up the first session and started the second. We determined that an error impacted the results of her first session, and the data values represented a single construct outlier. Since the data point cannot be corrected, we eliminated the observation of both sessions from the remaining analysis following best practices [AGJ13].

The Shapiro-Wilk test did not show evidence of non-normality in the self-reported confidence for the surveys ($W = .97$, $p = .387$) or intermedial Survey ($W = .95$, $p = .082$). With 20 participants conducting the study, we do lack some statistical power. However, as we are also concerned with inflated Type-I error rates, we adjust the p-values across separate testing models, following the Bonferroni correction method.

As seen in the upper chart of Figure 4.8 and upper-left chart of Figure 4.9, participants reported slightly higher confidence in the guided sessions than in the unguided sessions, except for the visualization survey scores. Users reported increased confidence in all cases compared to the self-competency control survey (solid line). Again, the visualization response scores are an exception. We see that participants reported less confidence on average in the guided session survey than in the control survey. The bottom chart shows that guidance positively impacts users' confidence when it is received in the first session.

The end-of-session survey results are consistent with results retrieved from the intermedial surveys. Participants responded to more questions in the guided sessions than in the unguided sessions, and also more in the first than in the second session. These results show a higher rate of decision reluctance in the first sessions and the guided sessions. The increase in confidence from the self-competency survey (Survey 0) is consistent with the farming survey scores. Participants felt less confident in their ability to make farming decisions compared to interpreting visualizations. The observations in our dataset are independent.

4.9.1 Global Characteristics

We will hereby refer to *network size* as *actions* and to *network diameter* as *time*. The sample's average time of all sessions ($n=38$) is 24.49 minutes, 21.81 minutes for the guided sessions ($n=19$), and 27.17 minutes for the unguided ($n=19$). Participants relied

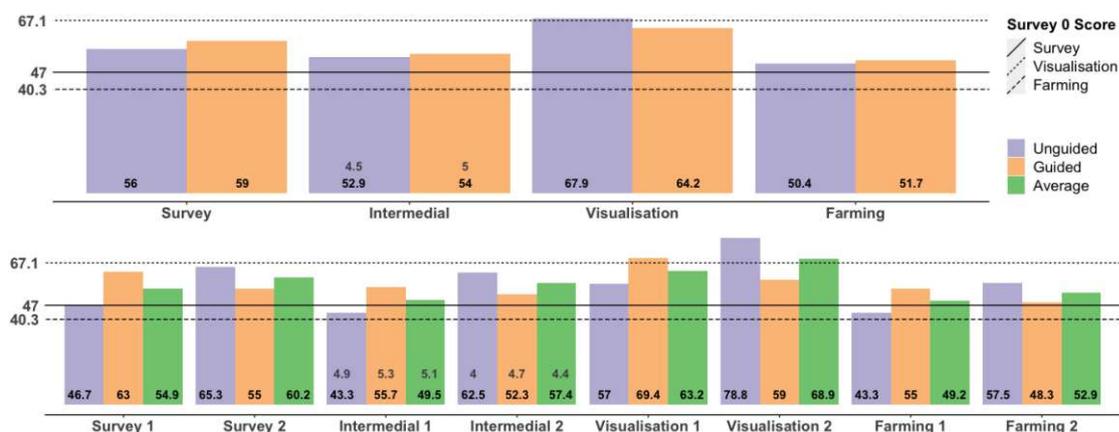


Figure 4.8: The upper bar chart outlines the overall average score for guided vs. unguided sessions in each self-reported category. The gray number above the intermedial survey scores denotes the average number of pop-up questions. The horizontal lines indicate the average scores for the control survey (Survey 0). The bottom bar chart outlines the self-reported scores per guidance and averages for each session. We further split it into two main question groups: confidence using visualizations (Visualisation) and confidence taking farming decisions (Farming). The number above the intermedial survey scores represents the average number of pop questions answered. Finally, labels 1 and 2 on the x-axis refer to the session number.

heavily on the analysis in cluster B, the crops information cluster. Only in cluster D, the decision cluster, is the time spent higher in the guided session. This is a result of more frequent changes in decisions when receiving guidance. Moreover, Figure 4.10 shows that, on average, participants took approximately 30 more actions in the unguided sessions to reach a decision, which indicates that the guidance helped accelerate the decision-making process.

The actions ($W = .92$, $p = .01$) and time ($W = .92$, $p = .01$) variables are not normally distributed. Thus, we conduct the Spearman rank correlation test to respond to hypothesis **H1a**. We test four hypotheses by dividing the alpha value by 4 following the Bonferroni correction method. Thus, the adjusted alpha-value is .013. There is a statistically significant negative correlation between the *intermedial survey* and each of actions ($r(36) = -.41$, $p = .01$) and time ($r(36) = -.41$, $p = .011$). The survey scores yield no statistically significant result with the global characteristic metrics proposed in section 4.7. We conducted a one-way ANOVA test to respond to hypothesis **H1b** and examine whether guidance or session impacts the proposed *global characteristic metrics*. We adjust the alpha value to .013 following the Bonferroni correction method for the four comparisons. The guidance does not significantly impact actions ($p = .224$) or time ($p = .176$). A one-way ANOVA demonstrates that the session order statistically affects the actions performed ($F(1, 36) = 7.6$, $p = .009$). We also find a statistically significant main effect on the time ($F(1, 36) = 19.38$, $p = 9.19e - 05$). Therefore, **H1a** is accepted, and **H1b** is not confirmed.

4. CONAN: MEASURING AND EVALUATING USER CONFIDENCE IN VISUAL DATA ANALYSIS UNDER UNCERTAINTY

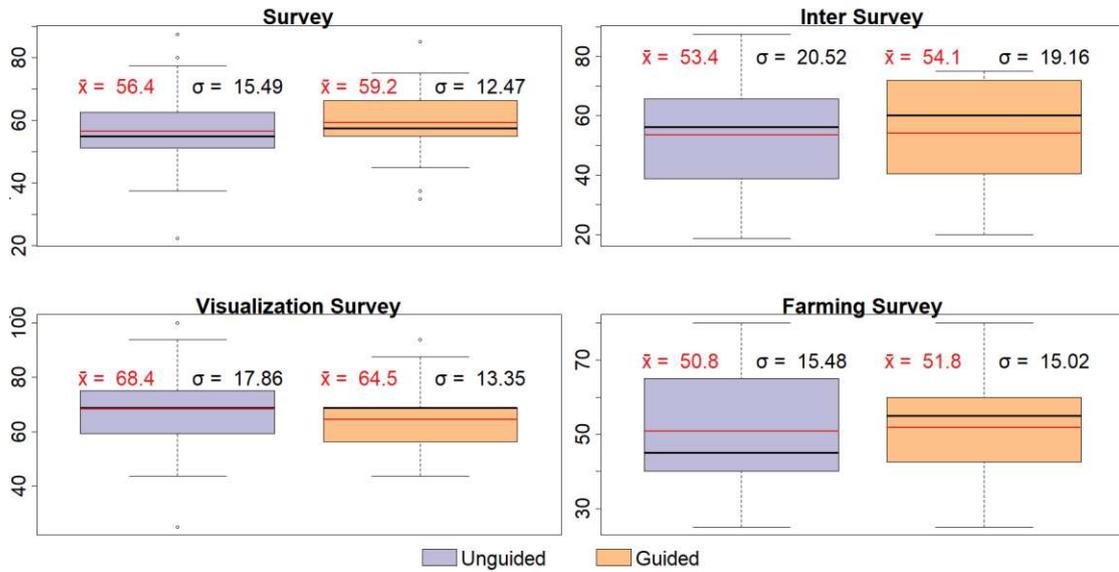


Figure 4.9: Distribution of the survey, intermedial survey, visualization survey, and farming survey scores for the guided vs. unguided sessions.

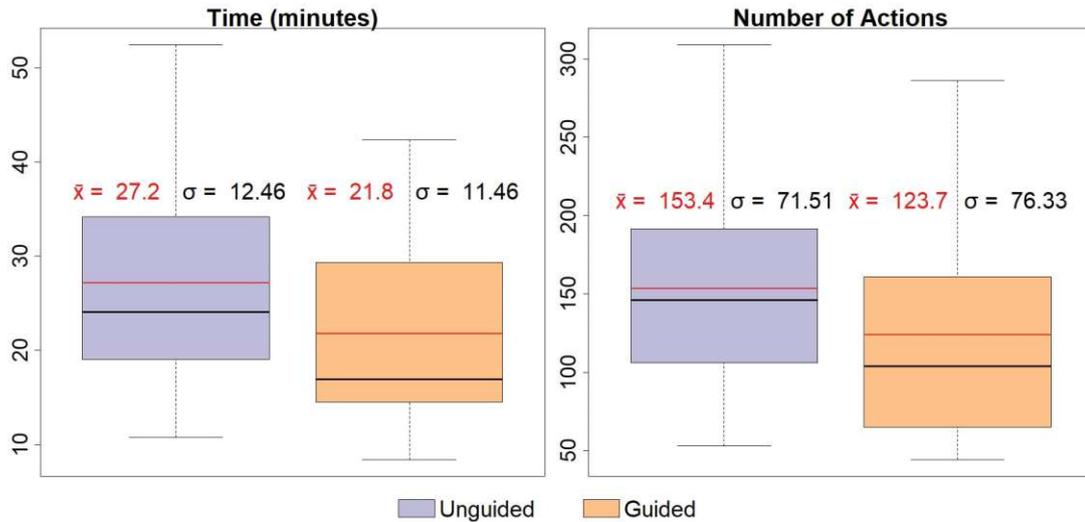


Figure 4.10: Distribution of time and action in each of the guided vs. unguided sessions.

4.9.2 Global Topology

The graphs' centralization ($W = .96, p = .191$) and density ($W = .96, p = .136$) variables did not show evidence of non-normality when we conducted the Shapiro-Wilk test. The average centralization of the network graphs is 0.334, and the average density is 0.156. In Figure 4.11, we can see that the graph's average centralization slightly increases in the guided session. Interestingly, Figure 4.11 shows that the centralization in the unguided

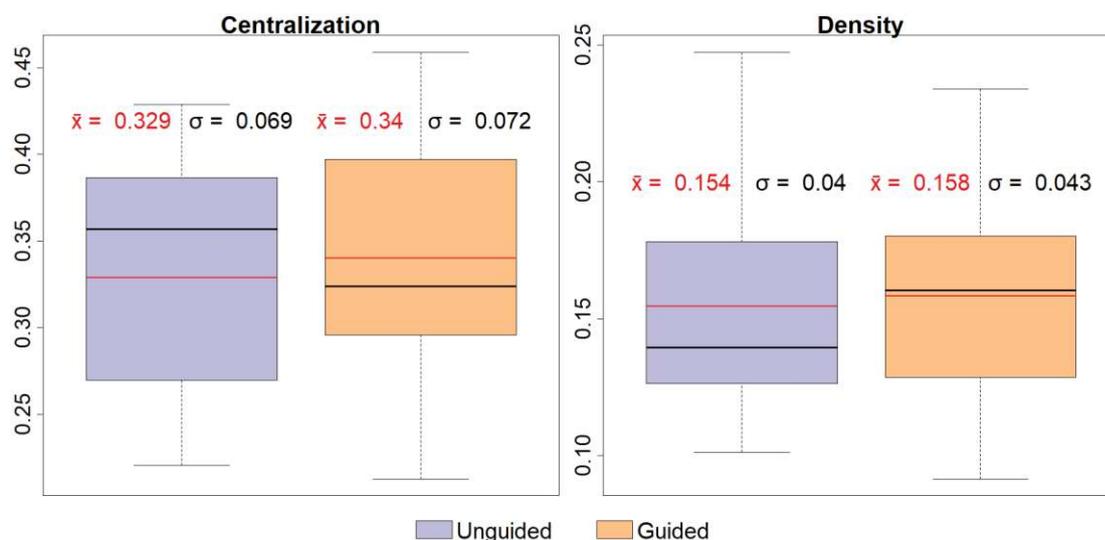


Figure 4.11: Distribution of the network's centralization and density metrics for the guided vs. unguided sessions.

session is negatively skewed. On the other hand, the centralization results are positively skewed in the guided sessions. This pattern is inverted in the density distribution. We test two hypotheses; thus, we adjust the alpha-value to .025. We run a multiple regression to test hypotheses **H2a** and **H2b**. We want to examine the prediction significance of the global topology metrics on the farming survey scores.

We used multiple linear regression to test if centralization and density (global topology metrics) significantly predicted the farming survey scores. The fitted regression model was: Farming survey score = $35.48 + 81.66 \times (\text{Centralisation}) - 73.37 \times (\text{Density})$. The overall regression was statistically significant ($R^2 = .18$, $F(2, 35) = 4.936$, $p = .013$). Centralisation significantly predicted farming survey scores ($\beta = 81.66$, $p = .018$). However, density did not significantly predict the farming survey scores ($\beta = -73.37$, $p = .202$). Then, we test hypothesis **H2c** by conducting a two-way ANOVA correlation test to investigate the impact of the guidance on the global topology metrics. The test revealed no statistically significant impact of the guidance on the global topology metrics ($F(1, 36) = .404$, $p = .529$). Therefore, **H2a** is accepted. **H2b** and **H2c** are not accepted.

4.9.3 Local Topology

The view switch in each cluster is the most prominent and important action in the process. This is normal and expected since these are actions needed to switch among the three primary tabs seen in Figure 4.2, Figure 4.3, and Figure 4.4. We removed these three actions from Figure 4.12 for a better analysis. We notice that the actions in cluster B (crop information tab) are the most prominent in the analytical process. However, that is not the case for betweenness centrality, where actions of cluster C (crop companion

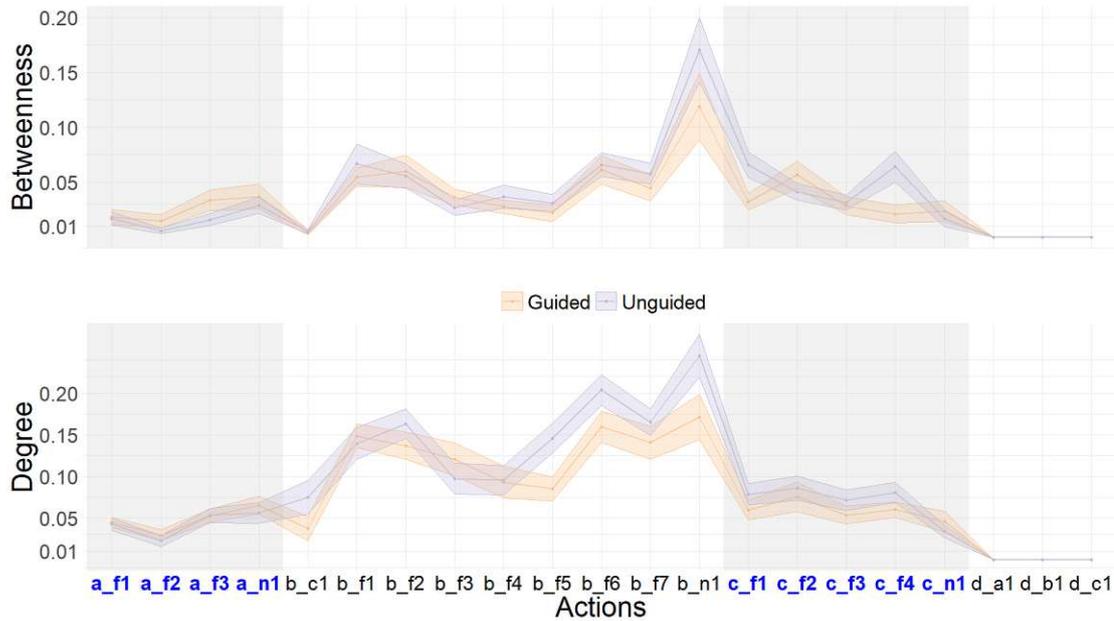


Figure 4.12: Line plots illustrating a comparison between average betweenness and degree centrality for each action in guided vs. unguided sessions. The gray boxes highlight the separate clusters. Among several findings, the plots highlight the potential importance of the Crop Information view (cluster b) to users’ decisions.

tab) compete for importance, particularly in unguided sessions. We can see that guided sessions negatively impact the participant’s reliance on cluster C, the crop companion view. This impact is more consistent in the measure of degree centrality. The actions performed in cluster C seem less prominent in the guided sessions. In the guided sessions, we can see an increase in users’ reliance on **b_f3** (filtering to the growing view) and, to a lesser extent, on **b_f1** (filtering to the farming calendar in the crop information tab). Actions **b_f5**, **b_f6**, and **b_f7**, distances, yields, and prices plots, respectively, become noticeably less prominent in the analytical process of the guided sessions. Also, as denoted by **b_n1**, taking notes in cluster B is less important and prominent to the participant’s analytical process in the guided sessions. The actions in the clusters are generally the least important and prominent to the participants sense making and decision making. They slightly increase in importance with the guided session.

4.10 Discussion

Our results, outlined in Figure 4.13, show that a network graph representation and metrics of analytic provenance support a better measurement of user confidence. We outline lessons learned and limitations of our study.

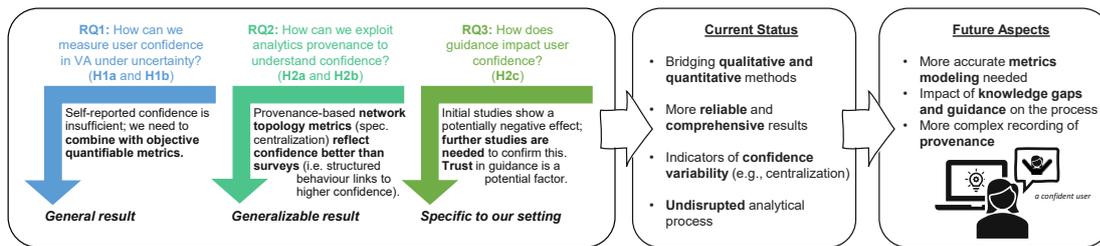


Figure 4.13: A visual summary of the aspects tackled through our research questions, take-home messages describing the current status, and future prospects.

4.10.1 Self-reported Confidence in Relation to Global Metrics

Our proposed hypotheses investigate the significance of the proposed metrics. The results demonstrate that self-reported confidence is only reliable with rigorous study design considerations. For example, we see in Figure 4.8 a discrepancy in sessions' self-reported confidence between the *intermedial survey* and the end-of-session *survey*. One must account for several factors when collecting self-reported confidence. The three main factors are the self-perception of confidence, the stage upon which the questions are asked, and the wording of the questions. The intermedial survey responses might be more accurate than the end-of-session surveys on a participant's level, as they quiz the participant immediately after making the decision. Interestingly, the intermedial survey scores reported in the first session more accurately reflect the farming decisions' self-reported confidence in the same session.

Furthermore, the intermedial survey is specific to a single task. Thus, the participants can provide a more accurate assessment of their self-confidence in the intermedial survey. Given that assertion, we lean to accept **H1a** as results show a moderate correlation between the global characteristic metrics and the self-reported confidence in the intermedial survey. While the correlation does not show with the survey results, we see that it may possibly support our claim that self-reported confidence is insufficient. This assumption is also based on the analysis of discrepancies seen in Figure 4.9 in the self-reported confidence between the intermedial and end-of-session surveys, Figure 4.9 (top chart).

Although our results show that the session impacts the participant's performance, we could not prove the effect of guidance on confidence. Reviewing the descriptive results of Figure 4.8, we conclude that guidance could potentially impact self-confidence. We can see that the confidence in self-reported surveys decreased in the second session when they received the guidance. This negative impact is interesting to explore. However, the sample size and study design must be increased to test the significance of the impact of guidance on the metrics. We also believe that a qualitative study of the results can complement the findings and shed light on the quantitative results. In our study, we monitored participants' analytical behavior and had a short discussion with them at the end of the sessions. One explanation for the results regarding the impact of guidance on the analytical process is that users who received guidance in the second session arrived

with a particular cognitive model of sense making and decision making. The guidance in the second session challenged their interpretation of the data, leading to a lack of confidence. The average number of intermedial survey questions answered in Figure 4.8 shows that participants change their decisions more often with the guided session.

DeMarree et al. [DPBn07] could provide a possible explanation of the aforementioned results for guidance. They affirm in their study that the inconsistency of information negatively impacts self-confidence. They also attribute a lack of confidence to unorganized knowledge. In our study, guidance provided in the second session may have contributed to inconsistent information, and it was not provided in an organized manner. Furthermore, the guided sessions presented repetitive guidance. According to DeMarree et al. [DPBn07], ruminative thoughts may undermine self-confidence. Finally, users tend to exaggerate their self-confidence when they do not have a basis for comparison. Once presented with guidance that contradicts their analysis, their self-confidence decreases.

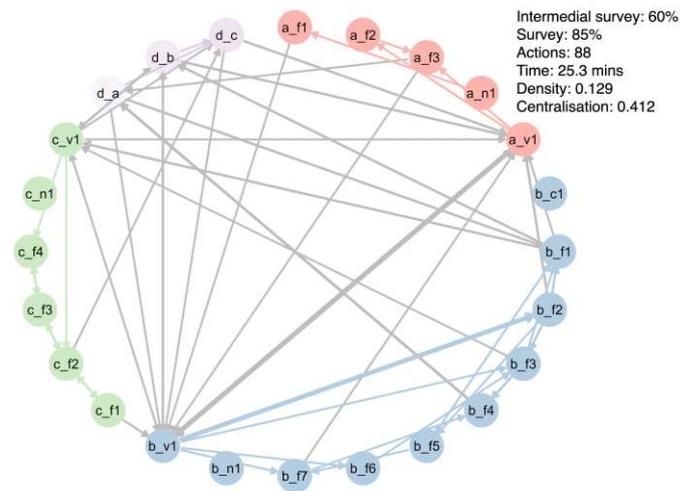
4.10.2 Lessons Learnt

The guidance received in the first session might have influenced the analytical behavior of users in the second session. This result could highlight the importance of providing explanations in combination with guidance. It is supported by previous research findings, which assert that guidance explanation supports the effectiveness of the guidance tool [PMCEA⁺22]. Furthermore, adding the combination of participants receiving guidance in both sessions and not receiving guidance in either session may be helpful to the results.

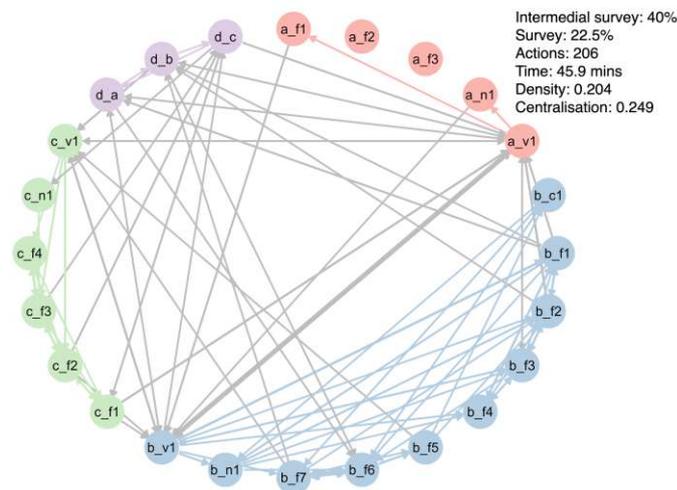
4.10.3 Confidence Reflected in the Network Graph Topology

As for the global topology metrics, we accept **H2a**, but we reject **H2b**. We specified the farming survey scores in **H2a** because we are interested in how well these metrics can predict the domain decision-making self-confidence. Centralization and density value could shed light on the users' behavior and self-confidence.

A further qualitative study of the network graphs illustrates the user's confidence level and highlights areas that contribute to possible low confidence. Two ProvNets are shown in Figure 4.14. We see a more structured and less dense analytical process in Figure 4.14 (a) compared to Figure 4.14 (b). P2 constantly jumps from one action to another, demonstrating confusion and low confidence. We see a heavy density of intra-cluster connections, particularly in cluster B (blue). With the more confident P1, the intra-cluster arcs are not dense. We notice a more organized and consistent analysis process within clusters. The analysis session of P1 is provided in the supplementary material for reference [Mus24a].



(a) P1: more structured, potentially more confidence



(b) P2: less structured, potentially less confidence

Figure 4.14: (a) A ProvNet of a guided first session (participant P1–55314466) that resembles the analysis of a participant who displayed high confidence across our metrics and the self-reported surveys. The opacity of the Cluster D nodes illustrates the percentage of decisions that followed the guidance. (b) A ProvNet of an unguided first session (participant P2–41277257) that resembles the analysis of a participant who displayed low confidence across our metrics and the self-reported surveys. The node labels in both networks are related to actions outlined in Table 4.2.

P2 occasionally decides on the companion crop after analyzing crop information in cluster B. A natural pattern should be making that decision after analyzing the companion data in cluster C. Nevertheless, they tend to follow their companion decision by analyzing the crop companion data for confirmation. We can see in Figure 4.14 (a) that the more confident participant **P1** makes their decision following **c_f2**, an analysis of crop companion benefit from repelling insects. Although in both ProvNets we see that the most frequent tie is from cluster B to cluster A, the weather analysis in cluster A does not provide enough confidence to decide on the planting week. P2 rarely filters the weather view, where **a_f2** and **a_f3** are not connected to the analytical process.

In our study, we learned that our proposed network representation can be suitable for examining user confidence during sense making. This phenomenon is *difficult to capture with self-reporting surveys*. Although the survey typically poses questions about the sense-making process, the participants' responses are influenced by the stage at which they respond. Posing questions to participants immediately after each sense-making action is not feasible, as it disrupts the process. The objectively quantifiable metrics can potentially complement self-reporting surveys in providing insight into users' confidence.

Furthermore, our findings indicate that self-reported confidence might be unreliable as a sole metric. We experienced that, instead, a combination of qualitative assessment with quantifiable centrality metrics can be more dependable indicators of confidence. The former offers insights into the sources of confidence variance, while the latter demonstrates how guidance influences user confidence in analytical processes. Although our results show that guidance indeed impacts user confidence (responding to **RQ1** and **RQ2**), further studies are required to answer **RQ3** comprehensively.

4.10.4 Externalizing and Validating Confidence Metrics

Future studies may explore the interplay between *confidence*, examined in our study, and *trust* as two interconnected properties of the user attitude. In our study, we did not need to filter out trust from the network graph because trust cannot be externalized with a network graph. Network graphs are suitable for externalizing confidence. In contrast, as explained in section 4.3, trust is the property of a user's attitude towards the system at a given time of an interactive session. Thus, to externalize trust, one needs to maintain the order of events and examine nodes at a lower level, i.e., the event level. Sequential or hierarchical tree graphs, for example, Stitz et al.'s [SLSG16] provenance graph, can be more suitable for the examination of trust. Measuring user self-confidence throughout the session and at each event requires a more rigorous approach, which we leave for future research.

In section 4.7, we rely on the literature to reason about the potential interpretations of the possible metric observation results. We incorporate a social psychology perspective in our analysis to examine the properties of a human attitude. In designing our study, we deployed a real-life task-based study, maintaining an uninterrupted and complete analytical process, and designing a dashboard in a domain usable by the general public.

We set concise hypotheses to analyze the results before the evaluation sessions, and we plan to examine further hypotheses in future studies. In our analysis, we compared the metrics to the self-reported surveys, which are a commonly used tool to measure user confidence. We accurately report the statistical significance of our correlation and prediction tests with promising results. However, the validation of the confidence metrics can be strengthened by further studies with a larger sample and qualitative analysis, a more accurate metric modeling, and a more complex recording of provenance. Finally, our work thoroughly reports the study procedure and archives the raw results and analysis code in our supplementary material [Mus24a].

4.10.5 Limitations and Bias

Without a readily available solution, building a new dashboard for the study required certain compromises. Our dashboard simplifies the interactions to limit potential bias from the user being unable to fully control the dashboard. The tracking technology limitation contributed to our design decisions, outlined in subsection 4.5.3. Allowing the user to interact freely with all dashboard components at all times could have provided more comprehensive provenance data.

Moreover, the size of our networks might have hindered a comprehensive examination of the metrics. The ideal network size for a topological analysis requires further studies. However, small networks can impede the topological analysis of a network [Xu11]. Thus, we were limited to a handful of metrics that could be helpful. Considering typical metrics in SNA, such as cliques, clustering coefficient, and modularity, could be valuable in future studies. We note several recommendations for expanding the metrics used to measure user confidence. The *ascendancy* metric [Ula00] to study expansion in ecosystems could be helpful in reverse to study the effectiveness of guidance on users' sense making and decision making. Moreover, *residual freedom* [Ula00] can also indicate an effective visualization that allows the user to uncover more insights than determined by the developer, who is often not a domain expert. A mature method to quantify the knowledge flow in the interactivity network could also support using the *Average Mutual Information* metric [Gon19] to quantify the inferred confidence.

Finally, in our approach, we pre-define actions that are possible for the users. However, this may have introduced bias in the study results. A more complex approach to recording user analytical actions, such as eye-tracking, could significantly improve the reliability of the proposed metrics. A more comprehensive qualitative examination of network graph structures and their relation to confidence is an interesting path for future studies. Subsequently, it is important to investigate the impact of users' expertise on their confidence in the context of a guidance-supported VA.

Summary We conceptualized user confidence and distinguished it from user trust. We then proposed SNA metrics to evaluate user confidence in VA under uncertainty

4. CONAN: MEASURING AND EVALUATING USER CONFIDENCE IN VISUAL DATA ANALYSIS UNDER UNCERTAINTY

and conducted a study to examine the value of the proposed metrics. The global characteristics, global topology, and local topology metrics offer the potential for an insightful analysis of user confidence in VA. They can complement the commonly used self-reporting surveys. An interesting direction for future work would be to explore the distinctions between confidence and trust—quantitatively and qualitatively—within the specific visual data analysis context to further understand their respective impacts on user interactions and decision-making processes. We also see potential in investigating explainable guidance to study the impact of explainability on user trust, which further influences their confidence in sense making and decision making. With better modeling and tuning, we anticipate that our proposed metrics could improve the assessment of user confidence and, consequently, support a better evaluation and development of VA tools, including guidance.

TrustME: A Context-aware Explainability Model to Promote User Trust in Guidance

This chapter addresses the need for an explainability approach to support guidance designers in addressing the issue of user trust in a guidance-supported VA system, and is based on the publication:

Maath Musleh, Renata G. Raidou, Davide Ceneda, “TrustME: A Context-Aware Explainability Model to Promote User Trust in Guidance”, in *IEEE Transactions on Visualization and Computer Graphics*, Volume 31, No.10, 2025, pp. 8040–8056, doi: 10.1109/TVCG.2025.3562929.

Guidance-enhanced approaches are used to support users in making sense of their data and overcoming challenging analytical scenarios. While recent literature underscores the value of guidance, a lack of clear explanations to motivate system interventions may still negatively impact guidance effectiveness. Hence, guidance-enhanced VA approaches require meticulous design, demanding contextual adjustments for developing appropriate explanations. In this chapter, we discuss the concept of *explainable guidance* (XG) and its impact on the trust relationship between the user and the system in VA. We, then, propose an XG model to design an effective explainability strategy to boost user trust in guidance. We demonstrate the usability of this framework through use cases. We also evaluate it in a design walk-through with a guidance expert to showcase how our model supports designers in clarifying the rationale behind system interventions and designing explainable guidance.

5.1 Introduction

Explanations in Visual Analytics (VA) can shed light on human- and system-made decisions and interventions throughout the analytical process. In particular, support from the system, namely, *guidance* [CGM⁺17b], requires appropriate motivations and explanations to enable understanding of complex data interactions and ensure clarity within the analytical process while also enhancing user trust. Even if the afforded guidance is technically sound, its adoption in several domains might still be limited if users fail to understand it [MMT⁺23], mainly due to a lack of *guidance explainability*. While explainability is not a new concept, for example, it has been investigated in artificial intelligence (AI), explainable guidance (XG) is a concept that has yet to receive sufficient attention.

Guidance in VA can support sense-making and decision-making processes by addressing challenging situations in which the user struggles to complete the analysis independently. In contrast to *onboarding* [SCW⁺22], which targets the initial process of integrating and acclimating new users to a system, guidance provides continuous task-specific support throughout the entire analytical process in several forms [CGM19b]. For example, if the user reaches an impasse during the analysis, the system can provide a recommendation list of the next possible steps that can produce useful insights. The ultimate goal is to address the knowledge gap effectively by orienting the user's perception, directing the exploration process, or prescribing specifications to support the visual analysis [CGM⁺17b]. Thus, guidance supports several VA goals, such as enhancing user understanding, facilitating effective exploration, promoting insight discovery, and improving decision making—all while boosting user confidence throughout decision making [HS23].

The effectiveness of guidance is highly context-dependent [CAA⁺20, SEAAC24]. Factors such as user and analysis context or domain complexity [HS23, SCEA23, MAK⁺23] impact the effectiveness of guidance. This is especially true in critical domains like healthcare [MMT⁺23], where clinicians predominantly rely on predefined, clinically-approved models and workflows to minimize undesirable outcomes for the patient. In such *critical scenarios*, introducing guidance requires accounting for resistance from domain experts accustomed to conventional sense-making processes [SRM⁺19, OLA19] or considering a potential preference for guidance that provides less uncertainty [MMT⁺23]. This is particularly important when users make complex decisions between plausible alternatives with marginal trade-offs [DIPJ22]. In these situations, users base their trust on several factors, such as past experiences, conventions, or communication transparency, underlining the user's necessity to develop *trust* in the presented guidance. Thus, when designing guidance for such scenarios, the designer should actively mitigate the impact of the user and domain contexts on guidance while fostering credibility and understanding.

Sacha et al. [SSK⁺16, p. 243] describe trust as “*a process of passing and accepting knowledge between human and machine*”. In attitude theory, trust is a property of attitude and is considered an evaluation of an object [DPBn07]. In our context, the object of this attitude is guidance, manifested as visualized cues, specifications, or options.

Explanations applied to guidance can positively affect the sense-making and decision-making process while ensuring the stability of the VA process and the reliability of its outcome [DPBn07, PSBn23]—ideally boosting user trust. It is important to note that trust is built over time as it is subject to changes through use [Mui87], while “*sustained distrust can only lead to further distrust*” [Gam88, p. 234]. Ceneda et al.’s recent study [CCEA⁺24] found that trustworthiness depends on the explainability of VA guidance, affirming that it can also avoid confusion. Their study also highlights that XG can improve user acceptance of guidance because it mitigates the gap between the user’s expectations and the system’s guidance. Thus, it can address resistance among experts using VA tools. Other studies found that customized explanatory approaches can relieve cognitive overload [BKD⁺23], improve mental representations [SGRn09], and promote efficient sense making [YZE⁺24]. Hoque et al. [HSE24] illustrate how explanations can shed light on AI-produced guidance and improve issues such as understanding, transparency, and accountability. Moreover, a recent study on user confidence under uncertainty by Musleh et al. [MCER25] suggests that XG can improve the user’s analytical behavior.

For these reasons, we propose a design model for building and presenting explainable guidance as an essential factor for supporting the user in interpreting and understanding guidance and, consequently, for establishing and sustaining the user’s trust toward suggestions presented by the VA system. *We contribute:* 1. a discussion of the concept of XG and the impact of explanations on the user–system trust relationship; 2. a novel model for designing context-aware explainability strategies for guidance in VA systems, considering all aspects that designers need to address when designing effective explanations; 3. use cases and a walk-through with an independent expert in guidance design to demonstrate how to use the proposed model for an effective explanation of guidance in VA.

5.2 Related Work

This section discusses challenges and opportunities in applying explanations to guidance-enhanced VA. It also surveys explainability in VA and the impact of explanations on trust.

5.2.1 Explainability in VA

Explainable artificial intelligence (XAI) Explainability is a very active topic in AI and VA that refers to methods to inform the user about the inner workings of AI algorithms to improve the user’s trust and acceptance [LRBB⁺23, AR23]. XAI serves three *goals*: helping users understand a process, boosting users’ trust, and supporting a system’s usability [GSC⁺19]. Mohseni et al. [MZR21] present an end-to-end framework mapping XAI methods to design goals and propose appropriate evaluation metrics based on three user categories: novices, domain experts, and AI experts.

5. TRUSTME: A CONTEXT-AWARE EXPLAINABILITY MODEL TO PROMOTE USER TRUST IN GUIDANCE

Approaches to explain the results of AI models comprise both model-agnostic and model-specific solutions [HZW21]. These methods also depend on knowing why, how, and what explanations are needed [Wel19]. Although XAI designers use machine learning to generate post-hoc explanations [DAB⁺23, HMKB23], progressive implementations are also proposed to partially analyze deep learning models in an interactive way [CL18].

As *features of a good explanation*, Confalonieri et al. [CCWB21] identify the adaptability to the user and the context of the analytical process. Vilone and Longo [VL21] identify four pillars guiding explanation generation in XAI: end-user, goals, content, and language. They emphasize the recipient, questions to be answered, content, and terminology. Preventing miscommunication between the user and the system has also been stressed in literature as an essential aspect of generating effective explanations [EAM22]. Samek and Müller [SM19] view the interpretation step after the explanation to be a complicating factor in achieving adequate explanations. Hence, they underscore that designers should prioritize interpretability when evaluating explainable systems. More generally, Mohseni et al. [MZR21] suggest that interpretability is connected to the reasoning side, such as of the algorithm, while explainability reflects the property of the interface side.

Explainable guidance (XG) The features of a good explanation, as proposed in the XAI literature, are potentially transferable to XG. Explanations of guidance do not aim to explain an underlying AI algorithm. Instead, they aim to explain the guidance output in relation to the resolution of analysis tasks. Hence, while an overlap may exist between explainable guidance and XAI methods, the role, position, and aim of guidance within the VA system must drive the explanation.

Unlike AI systems, the user is a central aspect of the decision-making process in guidance-enhanced VA. In addition, while AI refers to intelligent algorithms behind the screen, such as influencing how a visualization on the front end functions and appears, guidance manifests as a *separate intrusive or non-intrusive visual encoding*, indicating that the visual environment's context impacts explanations. Finally, in XG, the explanation generates a *third visual layer*, in addition to the main visualization and the supporting (visual) guidance. On the other hand, in XAI, explanations appear as a second visual layer where the AI model is an integral component of the primary visualization, and the explanation supports it.

In general, *research in explainable guidance* is scarce. Several scholars emphasize the importance of adding explanations to guidance-enhanced VA systems [MMT⁺23, CD19, CCEA⁺24]. Ye and Johnson [YJ95] studied the impact of explanations on the user attitude and the effectiveness of guidance. They found that an explanation can improve the user's attitude towards the system and the acceptance of guidance. Arnold et al. [ACC⁺06] conclude that the user's knowledge level influences the appropriate type and form of explanation.

Guidance and onboarding techniques are often confused and used interchangeably. However, the two concepts differ in their goals and methods. Specifically, Stoiber et

al. [SWG⁺21, SCW⁺22] present a helpful perspective on distinguishing onboarding from guidance techniques in visualization. According to them, onboarding techniques typically address a usability knowledge gap, for example, how to use specific functions, while guidance tackles knowledge gaps in VA reasoning, i.e., how to solve a task and get insights. Focusing on explainability, an XG approach aims to reduce misinterpretations or ambiguities of guidance suggestions by designing and introducing appropriate explanations. We refer to this as the *interpretability knowledge gap*.

Although guidance and onboarding approaches differ in goals and methods (as highlighted), the literature on *onboarding techniques* can offer a bridge to improve guidance explainability. In particular, Dhanoa et al. [DWH⁺22], building on Stoiber et al. [SPA23]’s work, propose a refined process model for onboarding, describing concrete guidelines for designing system onboarding that we can roughly transfer and adapt to designing explanations for guidance.

5.2.2 Trust-Aware Guidance Design

Sacha et al. [SSK⁺16] investigated trust building during knowledge generation in VA. They highlight the importance of the trustworthiness of decision aids to the user. In their proposed framework, Ceneda et al. [CAA⁺20] emphasized trustworthiness as a requirement for designing effective guidance. Sperrle et al. [SJB⁺21] proposed a guided-VA system that explains the suggestions to boost user trust. The literature highlights explainability’s role in building user trust in guided-VA [CAS⁺18, HS23]—yet, building trust in the VA process remains an open investigation.

Goals of effective explanations Muir’s [Mui87] early work discusses the criticality of trust in developing guidance tools within mixed-initiative systems. Muir states that building trust is a precursor to providing effective guidance, suggesting that guidance requires three conditions to be trusted by the user: *predictability*, *dependability*, and *consistency*. More recently, Shneiderman [Shn22] states that to make sure that humans can rely on and use them effectively, we need to ensure that the system’s decisions, i.e., recommendations or predictions, are reliable, safe, and trustworthy, meaning that they can be correctly interpreted and understood (and consequently trusted), and used by their human counterparts. Ceneda et al. [CCEA⁺24] also mention, among others, explanations as one of the criteria that make guidance effective.

User background Ceneda et al. [CGM19b] and Sacha et al. [SSK⁺16] propose that designers unlock the guidance algorithm’s black box, showing how guidance is produced. Explanations could increase users’ trust in the information provided by the guidance cues, boosting user confidence in their decisions. Muir [Mui87] suggests that vague explanations, or lack thereof, negatively impact users’ interpretation of guidance cues. Guidance-enhanced VA caters to users with varying expertise and backgrounds. Therefore, understanding how users interpret guidance cues is essential in building an effective explanation, but adapting it to diverse users is challenging.

Guidance characteristics The context in which guidance is provided is also crucial for designing good explanations. The literature classifies guidance approaches by defining the degree, direction, inference, and analysis objective [CGM19b]. Sperrle et al. [SCEA23] elaborate on these classifications, proposing a framework for guidance strategies. Other guidance approaches consider the accepted intrusiveness and required detail-orientation of the guidance [MMT⁺23]. Collins et al. [CAS⁺18] extend on guidance approaches to define the type of knowledge needed to provide guidance: prior knowledge, session-specific knowledge, or situation knowledge. In thinking of suitable explainability, it is essential to consider the type of knowledge gap that triggers the guidance.

Level of detail Chuang et al. [CRMH12] suggest that considering the level of added details is crucial to support user interpretation. Thus, the system should carefully consider the level of explanation detail provided to improve its interpretability. As a model-driven visualization, explainable guidance should incorporate a progressive detail disclosure from a high-level abstraction to a low-level detail on demand. The progressive disclosure can improve users' trust in the tool [CRMH12]. De Visser et al. [dVCFP14] proposed a methodology to construct system trust cues that could be integrated into the explanation construction model to improve its adaptation to the analysis process stages.

Timing Presenting guidance at the right moment is also crucial [CAA⁺20]: untimely guidance could negatively affect users' trust. Effective guidance responds to a specific knowledge gap, and so should its explanation. As the user might not be aware of the knowledge gap, the system should account for the issue of unconscious knowledge gaps in constructing the explanations to boost trust in the guidance [CAA⁺20]. Han and Schulz [HS23] propose that the system should communicate the trustworthiness of the guidance to the user. The communication is improved when the system can accurately identify the user's state of trust. Recently, van Elzen et al. [EAA⁺23] proposed a framework to help the user communicate the trust level to the system. This framework characterizes trust dynamics within a machine learning pipeline and can be simplified and calibrated to be usable in a guidance workflow.

5.3 Explainable Guidance and User Trust

In this section, we clarify terms and concepts that we use to build our XG model.

Interpretability vs. explainability The literature shows that providing effective explanations in VA is a task that remains challenging [WDC⁺23] and that drawing from AI research, *explainability*—and the distinct yet interconnected concept of *interpretability*—are prominent research directions. However, there has yet to be a consensus on the definition of both concepts [HKPC19], which are at times used interchangeably to refer to understandability [CPC19]. In other literature, interpretability is seen as a property of the user's action towards the information presented, while explainability describes the characteristic of the information presented to the user [MZR21, Jau22]. The lack

of consensus is attributed to the concepts' roots in several fields outside computer science [JS22].

To interpret a visual output means “*to make sense of it for oneself* [the user] *and to learn what it means to others* [the system]” [Bar06, p.803]. This definition indicates that an interpretable object assumes at least one understanding, although many are possible. *Interpretability* refers, thus, to the attribute of a system's visual output that can be ascribed a *plausible* meaning. Conversely, *explainability* entails the *comprehensive, rational, and practical* explanation of a system's visual output from a source to a target. For example, imagine a dataset of patients' health records. A simple tabular representation would not suffice to explain how these patients are clustered. A practical alternative is to perform dimensionality reduction and clustering, and show the result on a scatter plot. The scatter plot is still not necessarily *interpretable* by non-expert readers. The *explainability* of the scatter plot can be improved by encoding the resulting, for example, two clusters with distinct colors, for example, red and blue. This explanation makes the plot *interpretable* for most readers. However, it can also carry multiple interpretations. One explanation can match the intended meaning of two distinct patient clusters, while another can (mis)interpret the color coding, for example, interpreting blue as representing surviving patients and red as deceased. While this interpretation is plausible, it does not align with the intended meaning.

Significance of interpretability to explainability Explainability and interpretability are interconnected within the VA workflow and in the context of guidance. Looking at linguistics and semiotics literature, during the *interpretation* of an object, the user constructs the underlying meaning by contemplating the form, motive, model, inputs, and context of this object [Bar06]. After building an interpretation, the user shares it implicitly through their analytical actions, for example, by informing the remainder of the decision-making process and shaping the direction of their subsequent inquiries. Thus, understanding how users interpret the system's visual output is crucial for evaluating the effectiveness of explanations. Most importantly, the interpretation of guidance must be *accurate* and *precise* to support the user's analytical process. While visualizations are typically polysomic [GTKI11], i.e., they are open to more than one interpretation, a guidance-enhanced VA should allow only *one* interpretation of the guidance, i.e., its unique intended meaning. This quality is required to ensure the explanation is interpreted correctly, i.e., unambiguously, clearly, and reliably. In this way, the user and the system can effectively “understand” each other, enhancing the user's analytical process with trustworthy insights [Mil19, WHLS19].

Explainable guidance and user trust Informally, we refer to *explainable guidance* as guidance that can *practically* provide a *precise human-understandable meaning* to the guidance output. This concept includes providing visual cues and describing the specifications or options that align the meaning with the user's understanding. As a result, the system boosts the user's trust in the provided guidance by presenting adequate explanations. In other words, explanations mediate knowledge-sharing between the

system and the user to enhance user trust [Cas04]. In this work, *user trust* refers to the user’s degree of certainty in the guidance provided by the system, which can be externalized by several means, such as the user’s willingness to follow the provided guidance [EAA⁺23].

5.4 Research Questions

Our literature analysis indicates that explainability is critical to support user interpretability and boost user trust in the system. Despite similarities to XAI, the challenges in XG are unique as a result of distinct mandates in the approaches and questions investigated. Thus, designing effective explainability strategies for guidance requires unique solutions. For these reasons, we developed a five-component model for designing guidance explainability, inspired by XAI and borrowing from VA onboarding. This model addresses the challenges of guidance-enhanced VA to improve the effectiveness of guidance through better explanations. We investigate the following *research questions*:

RQ1 What are the characteristics of effective guidance explanations?

RQ2 How can context-aware explainability enhance users’ attitude in guided-VA?

In the upcoming sections, we propose that an effective guidance explanation should be interpretable, reliable, timely, and cognitively non-overloading (**RQ1**). Furthermore, an effective explanation design requires a model that responds to several interconnected questions that consider contextual information about the user, domain, and VA session. We support this design model by practically demonstrating XG within various analytical scenarios and a design walk-through with a guidance expert (**RQ2**). In these use cases, we demonstrate the usability and benefits of our proposed model and outline step-by-step guidelines for using our proposed model to design explainability strategies for guidance tools in VA. Specifically, in the design walk-through with the independent guidance expert, we employ these guidelines to support the expert in designing XG strategies for a guided-VA solution.

5.5 Methodology: A Model for Explainable Guidance

We followed a three-step method (depicted in Figure 5.1) to construct our proposed model for XG. Initially, we surveyed the literature for explanation models in VA and their characteristics to construct a model draft using a bottom-up approach (Step 1). We subsequently refine our model using a top-down approach, condensing terms and iteratively validating the model (Step 2). Finally, we ensure the applicability of our model by applying it to practical use-case scenarios (Step 3).

Step 1. Bottom-Up Approach—Deriving the Model from the Literature We conducted a thorough literature review to gather relevant work and build our model. We performed

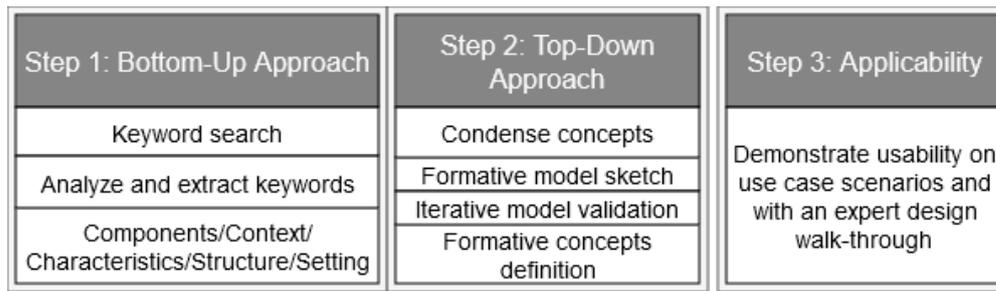


Figure 5.1: Three-step method to design the guidance explainability model.

query searches in Google Scholar using a combination of two sets of keywords. The first set includes variations of the term *explain*: *explainability*, *explanation*, and *explainable*. We combined these terms with a second set of keywords, including *model* and its scholarly interchangeable variations such as *framework* and *design*. We also included terms such as *characteristics* and its literary synonyms: *qualities*, *features*, and *notions*. For example, we queried ***allintitle: explainability model***, ***allintitle: explanation model***, and ***allintitle: explanation features***. We did not constrain the search to a specific time frame. By combining these terms and through the related queries, we generated a broad initial pool of papers spanning different fields, which we later scrutinized to evaluate their relevance.

We performed 21 different queries, and for each of them, we initially gathered and considered the first ten resulting papers. We further considered including more or fewer papers than the initial ten papers based on the query outcome and paper references. This process yielded an initial pool of 165 papers. We then reviewed the abstracts to discard any irrelevant works, including papers that studied the notion and design of explanations in the general computational field. For example, we excluded papers that discussed the topic of explanation to adolescents in schools. We considered all those papers that included a visualization and/or an analytical component relevant to our focus on VA. Then, the author of this dissertation reviewed the abstract and introduction of each section to select papers that discuss the design or characteristics of an explanation. We excluded non-peer-reviewed papers, duplicates, and workshop papers. Although workshop papers are valuable, we only selected journal papers, conference papers, and book chapters because we sought mature (and possibly validated) models. The clean-up resulted in 83 papers, including nine surveys.

The papers went through a subsequent in-depth analysis phase, which was initially pursued by the author of this dissertation, who manually extracted information on the proposed and inferred explainability components, as well as their most prominent characterizing concepts. First, these concepts were taken from survey papers. Then, proposed characteristics and components of explanation design that can directly or indirectly inform the XG model were progressively added from the remaining papers. To extract the concepts, the author of this dissertation reviewed all papers and added

the concepts, i.e., candidates for explanation characteristics, to an Excel sheet, coding its source paper and potential direct/indirect contribution to XG. Then, the author of this dissertation performed an initial revision of the gathered concepts and proposed their initial separation and grouping into separate sheets based on their potential role in the model. This activity was performed manually. Concepts and characteristics were then coded and grouped based on similarity to avoid *conceptual redundancy*, as illustrated in Step 2 below, through multiple iterations by multiple coders. This step yielded 406 concepts that directly or indirectly informed the explainability components, contexts, characteristics, and structure/setting. These are included in our supplementary material [Mus24b].

Step 2. Top-Down Approach—Refining the Model In the second phase, multiple coders used an inductive thematic analysis to construct a comprehensive model using the extracted concepts following a formative sketching approach. The author of this dissertation proposed an initial grouping of concepts based on the extracted components. Inspired by relevant literature, all authors of the published manuscript [MRC25] organized the concepts according to seven questions: who, which, why, what, how, when, and where [SCW⁺22, CLSdV22], which were then grouped into three main components (see Context, Structure and Setting in Figure 5.3). At this stage, we also considered and added to our model four factors that influence the structure of AI explanations: end-users, goals, contents, and language [VL21]. In an iterative process, all authors finalized the concepts and components relevant to our proposed model. We confronted the concepts against our experience in the field, developing and researching guidance for VA. To bridge, compare, and merge the concepts, we also comparatively analyzed the guidance features and evaluation proposed in the work of Ceneda et al. [CGM19b, CAA⁺20, CCEA⁺24].

To formalize *user expectations* towards guidance, we build our model upon existing literature on trust psychology and reflect on concepts from Muir’s framework [Mui87] on trust dynamics. In doing so, we account for diverse users and analysis contexts while ensuring that our model aligns with recent studies, suggesting that user trust is primarily affected by user expectations linked to the system’s past performance [OV21]. As part of our model’s epistemic foundation, we condensed the extracted concepts by identifying the most relevant *characteristics for effective guidance* [CCEA⁺24].

We employed several proposed approaches to evaluate and refine our model [KK17]. We tested the model against our experience and knowledge of the topic and applied it to well-known approaches and scenarios for further validation [MMT⁺23, CGM⁺18]. Eventually, we ended with an agile condensed model that can be employed in various scenarios (see section 5.6). In addition to the model, we also propose and provide definitions from the literature for the concepts used in the model to facilitate its understanding. We rigorously revised and polished these definitions to finalize a list of concise definitions of concepts, see Table 5.1.

Step 3. Ensure Applicability In the final step of the design, we demonstrate the usability of the model on various use-case scenarios taken from the literature and in a walk-through design scenario with a guidance designer. The demonstrations account for previous

user studies and participants' comments to infer issues that need to be addressed with appropriate explanations. In section 5.7, we demonstrate a designer's typical steps to use the proposed model effectively. We then present step-by-step guidelines to walk a guidance expert through designing explainability strategies for his guided-VA solution. During the walk-through, the expert reflects on the design process and XG strategy that he would follow, providing insights on the use of our proposed model.

5.6 Refining the Model: Guidance Explainability

We propose a model for designing guidance explanations, which requires designers to initially prioritize specific guidance **characteristics** while considering the **dynamics of user**

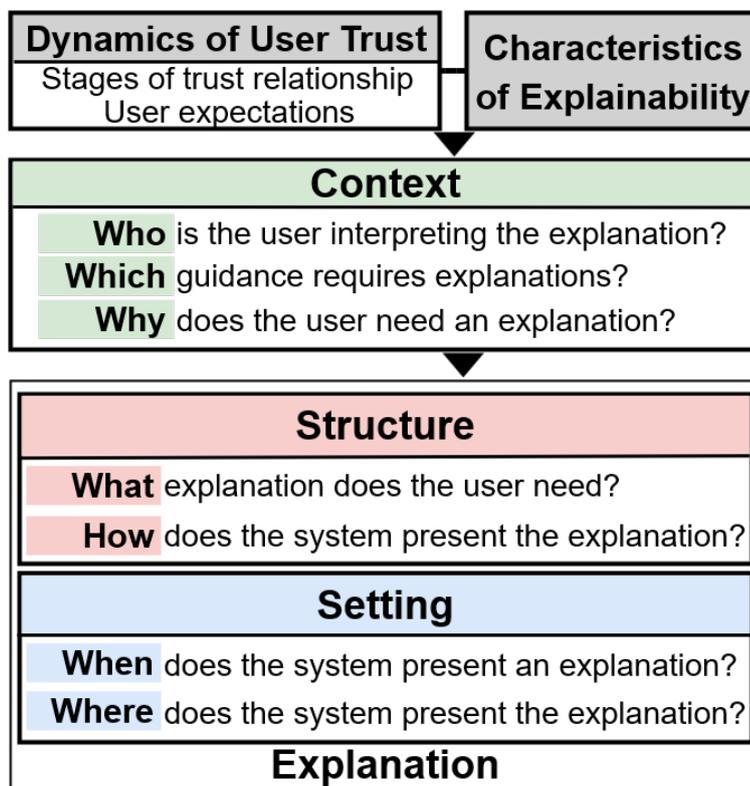


Figure 5.2: A model for designing explainable guidance. The **dynamics of user trust** (top-left gray box) inform the **characteristics** of the explanation (top-right gray box). Based on these two, **contextual** information drives the design of explanations by answering the questions “Who?”, “Which?” and “Why?” (green box). This information impacts the designer’s decisions on the explanation’s **structure** (red box) and **setting** (blue box). The questions are grouped by function, i.e., questions to design the structure (“What?” and “How?”) and questions to identify the setting (“When?” and “Where?”), forming together the **explanation**.

trust. The model also enables designers to focus on the specific **context** of guidance, in order to construct an adequate explanation **structure** and **setting** (see Figure 5.2).

5.6.1 Model Overview

Our model, depicted in Figure 5.2, is composed of multiple components that correspond to design choices, which affect each other and contribute to the overall efficacy of the explanations. The **dynamics of user trust** (top-left gray box of Figure 5.2) support the identification of the explanation’s characteristics by taking into account the dynamic user–system relationship. Moreover, it accounts for developing the user’s trust in the system and the guidance suggestions. In this regard, we consider the stages of the trust relationship and the user expectations towards the system. The dynamics of user trust inform the designer’s decisions in prioritizing specific **explainability characteristics** (top-right gray box of Figure 5.2) according to the different analysis phases, in which explanations are potentially needed. We derived a set of characteristics for the explanations in step two of our methodology, discussed in section 5.5.

When choosing appropriate explanation characteristics, the designer should provide an answer to three questions (green box of Figure 5.2: “Who?”, “Which?”, and “Why?”) to collect knowledge about the **context** in which explanations will be possibly needed and provided. The context informs the design of an explanation, taking into consideration two aspects: the selection of an appropriate **structure** for the explanation (red box of Figure 5.2) and the **setting** of the explanation (blue box of Figure 5.2). The former enables the designer to respond to “What?” and “How?” questions, and the latter answers the “When?” and “Where?” questions. The whole framework displayed in Figure 5.2 is an iterative process, as the setting of the **explanation** can impact structure decisions, and vice versa. An expanded version of the model, with the complete details of the different components, is shown in Figure 5.3. The upcoming sections discuss these components in detail.

5.6.2 ■ Explainability Characteristics

The literature describes numerous explanation characteristics [VL21]. However, depending on the application requirements, some characteristics are more critical than others to achieve specific results, for example, to improve guidance effectiveness. Prioritizing and choosing the most appropriate characteristics for the explanations can be challenging, as it often demands trade-offs [SF20] and the appropriate consideration of contextual information, which might not be entirely clear during the design phase. For instance, if designers aim to produce “simple” explanations, they might choose to compromise the completeness of the information conveyed in the explanation.

The matrix of trust dynamics in Figure 5.3 (top-left gray box) aims to direct this balance and help the designer decide which characteristics to prioritize. We discuss in detail trust dynamics in the next section. Given Muir’s [Mui87] trust dynamic framework and our study of guidance and its unique context, we conclude that an effective explainability

approach should consider four main *requirements*: 1) it should not overload the user’s cognitive process; 2) its annotation should be clearly distinguished from the guidance cues; 3) the explainability cues should integrate into the process when needed; and 4) it should not add more uncertainty to the sense-making process.

These four requirements inform the identification and the categorization of the characteristics displayed in the top-right gray box in Figure 5.3. Among these, we identify four primary characteristics, which we schematically organize as matrix elements. In particular, explanations should provide *cognitive relief*, be presented with *clarity*, be *timely*, and be *interpretable*. Based on the design requirements above, the designer may

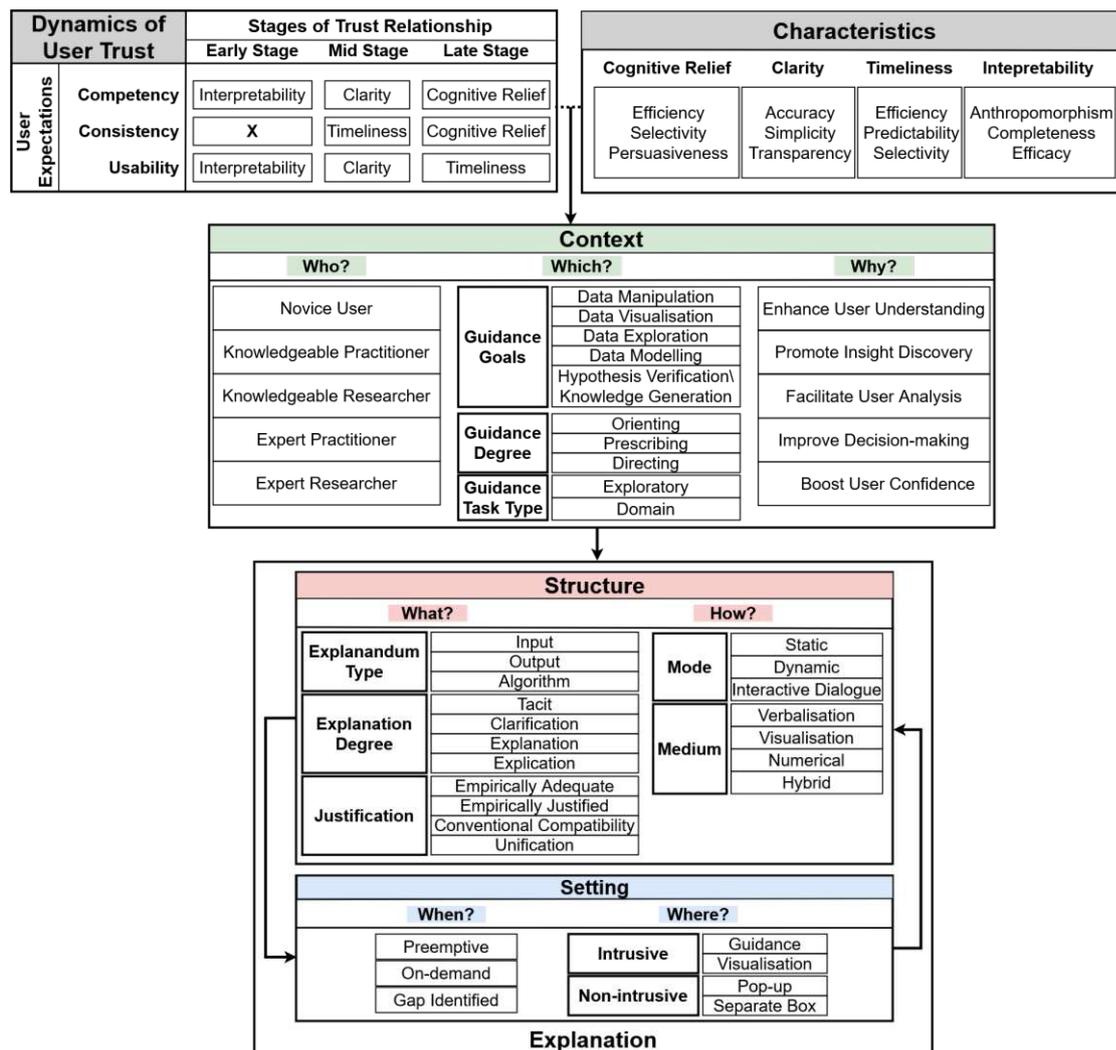


Figure 5.3: A detailed version of our proposed context-aware model to design explainability strategies for guidance. This schematic depiction is a detailed version of Figure 5.2, elaborating on all model sub-components.

need to prioritize a specific high-level characteristic, for example, clarity. These high-level characteristics are then fostered by secondary-level characteristics of explainability as outlined in Figure 5.3 (top-right gray box, see nested elements), for example, simplicity or transparency. In the subsequent paragraphs, we describe the four main high-level explanation characteristics and the secondary ones. For a complete description of these characteristics, we refer the reader to Table 5.1.

Cognitive relief refers to the decrease in the degree of dissonance between the explanandum and the user [VL21]. An explanation should reduce this dissonance and help align the system’s behavior with the user’s beliefs [CH97]. Cognitive relief can be achieved by prioritizing several secondary-level characteristics of explanations (shown in the top-right box in Figure 5.3) at design time, as follows: Trout et al. [Tro02] suggest that an explanation must be *efficient* to reduce the cognitive dissonance. Moreover, Yang et al. [YDH19] state that persuasive explanations help align the guidance behavior with the user’s beliefs. *Selectivity* further supports the *persuasiveness*, as demonstrated in previous studies [LF13]: the system should select and present only essential features of the guidance explanation.

Clarity refers to the quality of easily understood explanations, mostly related to the presentation of an explanation [BBV22]. Among the secondary-level characteristics that foster clarity, we mention: *Simplicity* is important to maximize *clarity* because simple explanations facilitate the user understanding [BBV22, SC20]. Barcellos et al. [BBV22] also suggest that *accuracy* reduces the ambiguity in information, facilitating understanding. Schoenherr et al. [SAM⁺23] propose that *transparency* makes a model visible and improves its clarity.

Timeliness refers to the quality of providing explanations when the user needs and expects them [CAGM22]. This characteristic is crucial to maintaining and boosting user trust [MVH⁺21, AKS21]. Timeliness is achieved by prioritizing specific explanation features. Mueller et al. [MVH⁺21] and Arnold et al. [AKS21] connect the necessity of presenting only relevant explanations to the user in an *efficient* manner with the timeliness of the explanation, demonstrating *selectivity* and efficiency [VL21, AKS21]. Finally, the explanation should remain *predictable* [AGN⁺20]. In other words, users should receive an explanation when they anticipate it.

Interpretability refers to the ability of the user to develop meaning from the explanation [Bar06]. While clarity measures the easiness of understanding, interpretability measures the *efficacy* of understanding [Bar06, BBV22]. This characteristic refers to the property of an explanation to produce the desired and intended results, for example, enabling the user to understand the intended meaning of the guidance. The designer should ensure that the system provides a *complete* explanation that the user could generalize to other cases [BDRD⁺20, SV23]. According to previous studies [Mil19, LS15],

explanations should be *anthropomorphic* to reproduce human characteristics, such as social behavior, and be more relatable.

5.6.3 ■ Dynamics of User Trust

Muir’s [Mui87] trust framework is a matrix that classifies the user–system relationship based on the user’s expectations vs. three different levels of experience. In our model, we consider a similar configuration and classify the user–system trust dynamics based on the *user’s expectations* vs three *stages of trust relationship*, as seen in Figure 5.3, top-left gray box. As the analysis progresses through its different stages, the user’s trust in the guidance evolves and changes, i.e., it might increase or decrease depending on several factors. The same happens to the user’s expectations from the guidance. Therefore, our matrix accounts for the evolution of trust and expectations in different stages of the analysis and proposes, accordingly, characteristics of the necessary explanations. In our matrix, we use the terms *early*, *mid*, and *late stage* to denote the three stages of the trust relationship. We also revise Muir’s three expectation levels that sufficiently describe the *user’s expectations* from the guidance: *competency*, *consistency*, and *usability*. These relate to explainability characteristics that the designer should prioritize in designing guidance.

User expectations The designer aims to meet the user’s expectations in maintaining and building trust in the guidance tool. Expecting *competency* refers to the user expectation that guidance demonstrates expert knowledge and technical facility [Mui87]. *Consistency* refers to guidance demonstrating repetitive behavior in its construction and presentation [Mui87]. Finally, *usability* refers to the guidance ability to help the user effectively perform a task [CCEA⁺24].

Stages of trust relationship Similarly to the user’s expectations, user trust develops differently according to the analysis stage [Mui87]. User expectations from system guidance appear in three stages of the analytical process: the early stage—also known as swift or default, the mid-stage, and the late stage [MZR21]. Distinguishing the three stages is a context-dependent problem. In other words, the designer must characterize the analysis regarding task workflow and possibly identify different analysis stages at design time. A common way to identify the analysis workflow and stages is by considering domain and analysis information, such as task descriptions [BM13]. For instance, an application that supports decision making in farming, for example, what and when to plant crops, might encompass three sub-decisions: choosing the main crop to plant, selecting companion crops, and deciding on the planting dates [MCER25]. In this case, the *early stage* of the trust relationship could be up until the user chooses the main crop. The *mid stage* could be up until the user selects a companion crop. Finally, the *late stage* lasts until the user decides on the planting date. Depending on the user context, the designer may also adjust the definition of stages. For instance, if users incrementally choose which crops to plant in sub-regions of their land, the stages can be defined based

5. TRUSTME: A CONTEXT-AWARE EXPLAINABILITY MODEL TO PROMOTE USER TRUST IN GUIDANCE

on the percentage of decisions taken. Thus, rather than defining the stage by the type of decision, we define it by the percentage of the analytical process completed.

Employing the dynamics of trust in guidance design To use the matrix, the designer should derive the primary user expectation towards the guidance when outlining the requirements at the beginning of the design process. In the *early stage* of the analysis, the user assesses the trustworthiness of guidance suggestions with little (or no) prior experience. In this phase, the user does not expect the guidance to be *consistent* because it is the initial encounter with the guidance, i.e., there is no basis for comparison. Instead, the user may expect a *competent* or a *usable* guidance. Thus, it makes sense for the designer to prioritize the *interpretable* quality of the explanations to avoid mistrust or alienating users at this stage.

In the *mid stage*, the user evaluates the guidance trustworthiness based on evidence from previous encounters. *Clarity* is most notably prioritized as a certain degree of trust has already been built between the user and the system. The shift from the focus on *interpretability* in the early stage to *clarity* in the *mid stage* happens because the user has already analytical experience that facilitates the interpretation of the explanation in the *mid stage*. When a user expects *consistency* from the guidance—for instance, when providing suggestions under similar circumstances—the designer should prioritize *timeliness* instead of *clarity*. This design decision aims to prevent delays that could cause confusion or mistrust. Ensuring consistency across similar situations takes precedence over providing the clearest possible guidance if it means sacrificing timeliness.

Finally, the user-guidance trust relationship matures at the *late stage*. If explanations have been correctly designed in previous analysis phases, the user has developed trust in the guidance's suggestions. In this late stage, efforts should be made to maintain and strengthen trust by *cognitively relieving* the user. In this sense, a system failure can damage the user's faith in its trustworthiness [Mui87]. Suppose the user primarily expects a *usable* guidance, for example, when the user needs the guidance to process data uncertainty. In that case, the designer should prioritize presenting a *timely* explanation when needed to avoid breaking the trust relationship.

5.6.4 Context-Awareness, Structure, and Setting

While the previous components focused on defining explanation features based on the trust relationship stages and user expectations, it is also essential to consider the context, structure, and setting in which explanations will be provided (Figure 5.3: green, red, blue boxes respectively). In our proposed model, we further elaborate on the explanation design by using seven questions inspired by the work of Stoiber et al. [SPA23] and Ceneda et al. [CGM19b] to consider context, structure, and setting in the design of explainable guidance.

■ Context

Explainable guidance needs to support sense-making and decision-making processes by addressing problematic situations and tackling the knowledge gap by orienting the user’s perception, directing the exploration, and prescribing specifications [CGM⁺17b]. To ensure that the guidance is accepted and used, the design of its explanation must consider several facets of the context (green box in Figure 5.2) in which explanations are provided: user experience and preferences, domain conventions and resistance, analysis session, and visual environment [SSK⁺16, SC20, BDRD⁺20]. Designing a customizable system that considers users’ preferences enables a more efficient and effective process, as variability in *users’ experience* has been documented to impact how users interpret explanations [LW90]. Furthermore, an explanation informed by the *domain context*, i.e., conventions and fundamental knowledge shared within a domain, is also anticipated to facilitate the acceptance of an explanation and overcome pitfalls for common domain-specific resistance [SSK⁺16]. Moreover, since users’ analysis strategies may differ, contextual *information about the analysis session* could be crucial to the explanation, while the *visual environment* is also set to influence an explanation’s visual appearance [SC20, BDRD⁺20]. For instance, explanations should be clearly visible within the interface and should not clash with the encoding of the rest of the environment. Understanding these four interrelated contexts informs the explanation design and how characteristics described in subsection 5.6.2 could be prioritized. For instance, based on the dynamics of the user trust matrix, if the designer aims to prioritize the cognitive relief characteristic, the system should not provide a beginner user with an elaborate explanation, as the user typically asks basic questions [LW90]. We summarize the three main questions about the context’s impact on the explanation’s design:

- *Who* is the user interpreting the explanation?
- *Which* guidance requires explanations?
- *Why* does the user need an explanation?

Who? Identifying the users’ context is crucial because they are the recipients of the explanations and need to interpret them. For instance, the user’s knowledge is crucial in a hypothesis verification process [SSK⁺16]. Most importantly, different users may require different kinds of explanations. We draw upon the XAI literature to further understand the intended users of an explanation. Barredo Arrieta et al. [BDRD⁺20] and Mohseini et al. [MZR21] classify XAI users by their intentions, i.e., developers or domain users. However, in typical VA sessions, guidance is usually provided to domain users who can be classified according to their knowledge gap and domain position, i.e., practitioners or researchers [MLC20]. A practitioner is primarily concerned with completing a task and making a decision. On the other hand, a researcher might be interested in acquiring more profound knowledge. Moreover, we can distinguish users according to their expertise, i.e., if they are novice, knowledgeable, or expert users [LW90]. As a consequence, our combined classification considers five distinct user categories: *novice*

5. TRUSTME: A CONTEXT-AWARE EXPLAINABILITY MODEL TO PROMOTE USER TRUST IN GUIDANCE

users, knowledgeable practitioners, knowledgeable researchers, expert practitioners, and expert researchers. Novice users would have similar requirements and expectations; thus, we do not distinguish them further.

Which? To develop adequate explanations, the designer must also understand the features of the guidance. In VA, guidance is designated by its goals, degree, and task type. System guidance can support several *goals*, including data manipulation, visualization, modeling, exploration, hypothesis verification, and knowledge generation [CGM19b]. These goals are anticipated to impact the structure of the explanation. For example, suppose the guidance goal in a VA system is to support hypothesis verification. The system should strengthen the user’s knowledge with a complete and accurate explanation of the guidance to allow the user to discover more insights.

Furthermore, guidance comprises three *degrees*—orienting, prescribing, and directing [CGM⁺17b]. An orienting guidance aims to steer the users’ analysis, especially when confused, lost, or lacking confidence. Directing guidance presents the user with several possible paths that facilitate an analytical process at an impasse. Finally, prescribing guidance suggests a specific direction for the user’s analysis. The *guidance degree* is expected to impact the designer’s decision on the appropriate explanation form. For instance, visual referencing might be more suitable to explain prescribing guidance than directing guidance because the latter is connected to several components in the visualization. Also, using highlights to map the explanation to the appropriate explanandum, i.e., the object of the explanation, might be a better solution than overloading the dashboard with different color encodings [RLN07].

Finally, guidance is designed to support either domain or exploratory *tasks* [CGM19b]. This design decision interwinds with understanding the user and domain contexts. For example, guidance that supports a practitioner in a clinical domain task requires an adequately correct explanation of the algorithms because the practitioner needs a “true” explanation to be confident in their decision.

Why? This question addresses the objective of the explanation and refers to the particular analysis and domain context that requires an explanation. This context is a crucial aspect of the explanation design, as the designer needs to reason about and, possibly, anticipate particular bottlenecks in the analytical process that could require explanations. When unexpected issues arise during a session, the system should explain the guidance to alleviate the issue. Explicitly identifying the objective of the explanation also informs a comprehensive context that enables the most appropriate decision on an explanation’s structure. Furthermore, identifying the objective of the explanation is important when evaluating the effectiveness during the evaluation phase of the system [MZR21]. We derive a set of low-level guidance explainability objectives from the literature of XAI [VL21, BDRD⁺20] and guidance [CAA⁺20]. These include *enhancing user understanding, promoting insight discovery, facilitating user analysis, improving decision making, or boosting user confidence*, as seen in the green box in

Figure 5.3. For example, if users require an explanation in the late stage of an analysis, they prioritize cognitive relief as they expect the guidance to be consistent. In that case, the explanation aims to boost the user’s confidence in their analysis or decision.

■ Structure

The context questions and the primary characteristics described so far inform the design of the form and content of the explanations. Next, we suggest thinking about how to structure the explanation. While context and characteristics define the high-level features of an explanation, the structure deals with organizing this information into a “physical” form. The structure of an explanation is, thus, related to the following questions:

- *What* explanation does the user need?
- *How* does the system present the explanation?

What? When designing explanations, designers should consider what content to include or exclude from the explanation. As shown in Figure 5.3, these choices depend on the demands of the context. Here, we need to look into three aspects: the *explanandum type*, the *explanation degree*, and the *justification* of the explanation.

The *explanandum type* considers the content of the explanation. We identify three types of explananda: the output, the input, or the algorithm of the guidance. An explanation may present the cause for the guidance by referencing its input(s). For instance, an explanation can refer to the input to show how it influenced or led to providing specific guidance cues. Alternatively, an explanation can reference the output(s) to explain the meaning of the guidance. For instance, a knowledgeable researcher may need to understand how each option will impact the analysis process. Finally, the context may require explaining the process or the algorithm that resulted in the guidance. For instance, an explanation of a guidance tool in a stock market VA system may need to explain the financial calculations to an expert but not to a novice user.

The *explanation degree* refers to the extent of the explanation. In some cases, users may understand the guidance without an explanation [Kha13], and designers should consider if explanations are indeed needed to support the goal of providing trustworthy guidance. As VA designers seek to make the visual environment and its guidance self-explanatory, a *tacit* [Kha13] approach refers to the intrinsic explainability property of the guidance object without adding any explanation layer, such as a textual note accompanying the guidance. The absence of an explicit explanation should not harm the effectiveness of the guidance tool. In all the other cases where an explicit explanation is required, the system can provide one of three degrees of explanation: a *clarification*, an *explanation*, or an *explication* [Wea64, Bau19]. Clarifications refer to explanations that make certain aspects of the guidance to the user without necessarily delving into a complete explanation of how or why the guidance was created. For example, the explanation can clarify to the user that a highlight on a slider refers to a precise area of interest. On the other hand, an explanation describes the complete rationale of how and why. These types of explanations

5. TRUSTME: A CONTEXT-AWARE EXPLAINABILITY MODEL TO PROMOTE USER TRUST IN GUIDANCE

provide a complete justification of the guidance based on the explanandum type. An explication instead illustrates the importance of guidance to the analytical process and its outputs. For instance, with an explication, the explanation can explain to the user how the guidance direction to analyze the vegetable prices at various points in time can impact their decision. The degree of explanation the user requires is expected to depend on designing a context-aware explainability approach. For example, guidance can demand a more profound explanation than guidance supporting an exploratory task.

The *justification* of the explanation refers to its level of being warranted. A designer could decide to provide a potential or rationally acceptable explanation for such suggestions without entirely verifying the truthfulness of guidance [Cup80]. When looking at the guidance literature, verifiability has not been documented as a typical requirement for effective guidance [CGM19b]. Based on Cupples’s discussion of explanations [Cup80], the designer should consider four possible justification levels. The highest level of justification is when the system presents an *empirically adequate*, i.e., true and accurate explanation. If this is not possible or necessary, the system may only present an *empirically justified* explanation based on accepted scientific knowledge without proving its truthfulness. The system may alternatively present *conventionally compatible* explanations, i.e., explanations that rely not on empirical evidence or justifications but on accepted conventions. Finally, in the absence of evidence or conventions, the fourth level of justification is a *unifying* explanation that shows how a case requiring guidance is similar to another case [Kha13]. This explanation could use an analogy or a direct comparison/contrast to justify the guidance.

To explain the differences between the justification levels, suppose that the guidance of a farming system suggests planting beans with corn [MCER25]. In that case, an empirically adequate explanation will elaborate scientifically on the fact that bean releases nitrogen into the soil and the positive impact of nitrogen on corn’s yield. An empirically justified explanation will refer to the accepted knowledge that legume companion crop improves corn’s yield. A conventionally compatible explanation will refer to the conventional knowledge of how these sister plants benefit each other without elaborating on scientific facts. Finally, a unifying explanation will devise an analogy that helps the user understand the positive impact of beans on corn crops.

How? At this point, the designer must decide the visual presentation of the explanation within the dashboard. As discussed above, the designer must consider the encoding employed within the visual environment to ensure that explanations are visible—without contrasting with or occluding the rest of the elements [RLN07]. For example, the designer could avoid using color as a visual channel if the rest of the visual environment is color-dense to avoid overwhelming the system’s users and undermining their trust in the guidance [DPBn07]. Drawing upon prior work [Mil19], we see that the two most important aspects of the visual presentation of the explanation are its mode and medium.

The explanation *mode* (shown in the red box in Figure 5.3) can be static, dynamic, or an interactive dialogue [Mil19]. Interactivity makes explanations more anthropomorphic,

as human explanations are inherently social [Mil19]. Imagine, for example, a dialogue between the system and the user where the former engages the latter by allowing them to request more information to form an explanation. Such setups are possible, for instance, with a natural language processing explanation model. On the other hand, a dynamic explanation adapts or changes based on the user’s interaction, but without direct conversation between the user and the system. For instance, a weather app adjusts forecast explanations based on user location changes. Finally, a static explanation could be more suitable for complex cases where the analytical process should not be interrupted—for example, a physician deciding on a therapy plan while contemplating several uncertainties.

The explanation *medium* (shown in the red box in Figure 5.3) is the channel through which an explanation is delivered. For example, in a stock market analysis visualization, the system may offer guidance to suggest what portion of the data the user should explore next. Drawing upon previous literature [EAJK⁺19, VL21], we see that explanation could take different forms; it could be: 1) encoded as a color reference that indicates the input that caused the guidance (visual explanation), 2) taking the form of a verbal explanation of the guidance rationale (verbalization), 3) a numerical formula that demonstrates calculations at the base of the guidance (numerical explanation), or 4) a hybrid explanation delivery mediums.

■ Setting

Finally, the designer decides where and when to provide an explanation, i.e., the setting of an explanation (blue box in Figure 5.3). The explanation’s context and structure impact the setting of time and space. In return, the setting influences the structure of the explanation, as indicated by the arrows between the red and blue boxes in Figure 5.3. For instance, the position of the explanation impacts the designer’s decision on a possibly suitable degree of explanation.

■ *When* does the system present an explanation?

■ *Where* does the system present an explanation?

When? The explainability approach can develop a preemptively transparent XG model [GMYA18], explaining to the user before an explanation is demanded or the system detects an actual trust issue. The designer identifies the need for such an explainability approach during the system requirement analysis in the early stages of the system development. At this point, the designer should identify possible bottlenecks that demand the incorporation of an explainability layer at certain stages of the analysis or for certain users. A preemptive explainability approach should be, at most, a clarification. Otherwise, it may disrupt the user’s trust in the guidance. Externalizing and communicating the user’s trust in the system is crucial for effectively designing guidance explanations [EAA⁺23], as an explanation can be offered instead when a specific knowledge or trust gap is identified. Detecting an analytical impasse, lack of confidence, or decrease in trust during the

5. TRUSTME: A CONTEXT-AWARE EXPLAINABILITY MODEL TO PROMOTE USER TRUST IN GUIDANCE

analysis process could trigger an explanation. For example, the system could provide explanations to boost users' trust in the guidance if a user ignores the guidance often. An explanation can be alternatively offered on-demand [GCV⁺21]. It is provided upon the user's request if they are unsure how to interpret the guidance, and it should be explained with completeness.

Where? Similar to guidance [MMT⁺23], explanations can be either *intrusive* or *non-intrusive*. In an intrusive explanation, the system interferes with the user's visual space by delivering the explanation or part of it, either on the visualization or the guidance itself, for example, by using color referencing on a visualization. An intrusive explanation should not construct an explication to avoid overloading the user and the visual space, which can be challenging. On the other hand, a non-intrusive explanation could be presented either in a pop-up or a dedicated field, usually in the margin of the visual environment. Complex explanations should be provided in a separate space (rather than intrusively) to avoid overloading.

Table 5.1: List of definitions of the concepts employed within our setting and context-aware model for guidance explanations. The colors link back to the components of our model, namely the dynamics of user trust and characteristics (grey), the context (green), the structure (red), and the setting (blue) of an explanation.

Basic concepts, appearing in section 5.3	Description
Explanation	A medium that mediates knowledge-sharing between the system and the user [Cas04].
Explainable guidance (XG)	A form of guidance that provides a <i>precise, human-understandable meaning</i> to the guidance output.
Explainability	The <i>comprehensive, rational, and practical</i> explanation of a system's visual output (or user's input, if seen from the user perspective) from a source to a target.
Interpretability	The attribute of a system's visual output (or user's input, if seen from the user perspective) that can be interpreted and given a <i>plausible</i> meaning.
User trust	The user's degree of certainty in the guidance provided by the system.
■ Dynamics of user trust	
Competency	The quality of guidance demonstrates expert knowledge and technical facility [Mui87].

Consistency	The quality of guidance that demonstrates repetitive behavior in its construction and presentation [Mui87].
Usability	The guidance ability to help the user effectively perform a task [CCEA ⁺ 24].
Early stage	The trust stage in the analysis process when a user assesses the trustworthiness of guidance suggestions with little (or no) prior experience—also known as swift or default stage [Mui87].
Mid stage	The trust stage in the analysis process when the user assesses the guidance trustworthiness based on evidence from previous encounters [Mui87].
Late stage	The trust stage in the analysis process when the user–system trust relationship matures with a user’s faith in the trustworthiness of the guidance [Mui87].
■ Explanation characteristics	
Cognitive relief	The degree of dissonance between the explanandum and the user [VL21].
Efficiency	The quality of the explanation to support the user reach a decision faster [VL21].
Selectivity	The quality of selecting and presenting only essential features of the guidance explanation [LF13].
Persuasiveness	The explanation’s ability to persuade the user to believe something [YDH19].
Predictability	The quality of user anticipation of when the system presents an explanation. [VL21, AGN ⁺ 20].
Clarity	The quality of easily understood explanations. It mostly relates to the presentation of an explanation [BBV22].
Accuracy	The capacity of an explanation to deliver true conclusion [BBV22].
Simplicity	The quality of an explanation to be constructed with simple elements [SC20].
Transparency	The quality of an explanation to allow users to inspect the elements that construct the guidance [SAM ⁺ 23].
Timeliness	The quality of providing explanations when the user needs and expects them [CAGM22].
Interpretability	The ability of the user to develop meaning from the explanation [Bar06].
Anthropomorphism	The quality to reproduce human characteristics, such as social behavior, and be more relatable [Mil19].

5. TRUSTME: A CONTEXT-AWARE EXPLAINABILITY MODEL TO PROMOTE USER TRUST IN GUIDANCE

Completeness	The extent to which the explanandum is explained [VL21]. The ability of explanations, under a controlled environment, to support users to make good decisions [VL21].
Efficacy	
■ Context	
Novice user	A beginner user who lacks experience in the domain.
Knowledgeable practitioner	A user with accrued practicing knowledge in the domain.
Knowledgeable researcher	A user with accrued research knowledge in the domain.
Expert practitioner	A user with high-level of expertise who is a leader as a practitioner in the domain.
Expert researcher	A user with high-level of expertise who is a leader as a researcher in the domain.
Guidance goals	The reason why the system provides guidance to the user [CGM19b].
Guidance degree	The level of guidance provided to the user: orienting, directing, or prescribing [CGM19b].
Guidance task type	The type of knowledge gap tackled by the guidance: domain task or exploratory task [CGM19b].
■ Structure	
Explanandum type	The guidance component being explained: input, output, or algorithm.
Explanation degree	The level of explanation provided to the user: tacit, clarification, explanation, or explication.
Justification	The degree of truth validation presented by the explanation: empirically adequate, empirically justified, conventional compatibility, or unification.
Explanation mode	The constancy of the explanation form: static, dynamic, or interactive dialogue.
Explanation medium	The form through which an explanation is communicated: verbal, visual, numerical, or hybrid.
■ Setting	
Preemptive	An explanation provided to users to forestall a possible mistrust issue.

On-demand	An explanation provided to users when they demand it directly.
Gap-identified	An explanation provided to users when system algorithm identifies possible issue.
Intrusive	An explanation that interjects the user's sense-making process by presenting it: on guidance, on visualization
Non-intrusive	An explanation that is presented as a separate, auxiliary view: pop-up or separate box.

5.7 Ensure Applicability: Designing Explanations

We demonstrate the practical use of our design framework by applying it to two analytical scenarios. The scenarios come from works discussing guidance-supported VA approaches, whose evaluation highlighted weak points in the respective guidance designs that could be addressed through appropriate explanations. Using our model, we describe how appropriate explanations could have been designed and structured to support guidance more effectively.

Finally, to validate the applicability of our design, we present step-by-step guidelines for applying the explainability model. We use these guidelines in a design walk-through with an expert in guided VA who has not been involved in this work. The session was 1 hour long and started with an introduction to our model. Then, we walked him through the steps of designing explainability strategies for the guidance tools in his previously developed VA solution for image selection in time and space [PMCM24]. In subsection 5.7.3, we describe how the designer follows the guidelines and assesses alternatives to design an explanation that can be integrated into his guidance.

5.7.1 Use-Case Scenario 1: Guidance in Eco-Farming

The first case is based on a guidance-enhanced eco-farming VA system in chapter 4. The work evaluates the impact of the guidance on user confidence in farming decision making (see Figure 5.5). The ■ *prescribing* guidance offers support to make an ecological farming decision making. The resulting guidance appears in a separate box at the bottom right of the dashboard (pointed in red). As concluded from a user study conducted with 20 participants, users need to receive additional support to understand how the system infers the recommendations. Taking the role of designers, we show how we would have designed the explanation a posteriori, which, in this case, was derived from the user study.

We adopt our context-aware model from Figure 5.3. Figure 5.4 depicts the resulting explanation design. The designer starts their work by identifying the *stage of the user's trust* based on the domain and application. In this case, we define the stages by completing each sub-decision: choosing the main crop to plant, selecting companion

5. TRUSTME: A CONTEXT-AWARE EXPLAINABILITY MODEL TO PROMOTE USER TRUST IN GUIDANCE

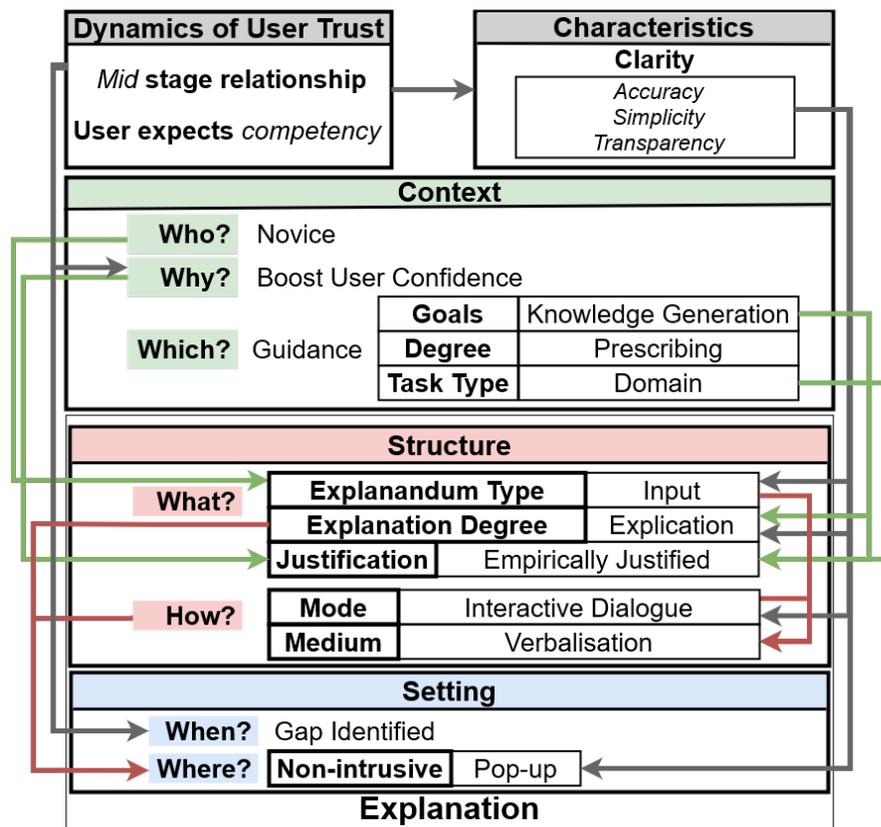


Figure 5.4: Reasoning flow of a designer for the first use case. The designer designs explanations for guidance in farming VA following our context-aware model (Figure 5.3). Decisions might be interlinked, as indicated by the arrows.

crops, and deciding the planting dates. Subsequently, the system tracks user confidence and trust in real-time. In this scenario, the user is in their **mid stage** of the analysis. The user has already chosen the main crops. During this stage, the user might have several **expectations** (see Figure 5.3); among which the designer prioritizes *competency* (Figure 5.4, *Dynamics of User Trust*). Although this expectation might change in later stages of the analysis, for example, the user may rather expect consistency later, and the explanation algorithm should be able to adapt to this change of expectations, we do not consider this case further within this use-case scenario for the sake of simplicity. Based on Figure 5.3, when prioritizing *competency* in the *mid stage* of the analysis, the designer should strive for **clarity**. It means that the system should deliver an explanation characterized by *accuracy*, *simplicity*, and *transparency* (Figure 5.4, *Characteristics*). We explain how the designer can pursue these characteristics below.

The system moves to identify the **context** in which explanations will be provided. In this scenario, the user study participant is a **novice** user with minimal background in farming or gardening (Figure 5.4, *Who*). Furthermore, during the study, the system detected hesitation from the user to follow the guidance, indicating a lack of trust and

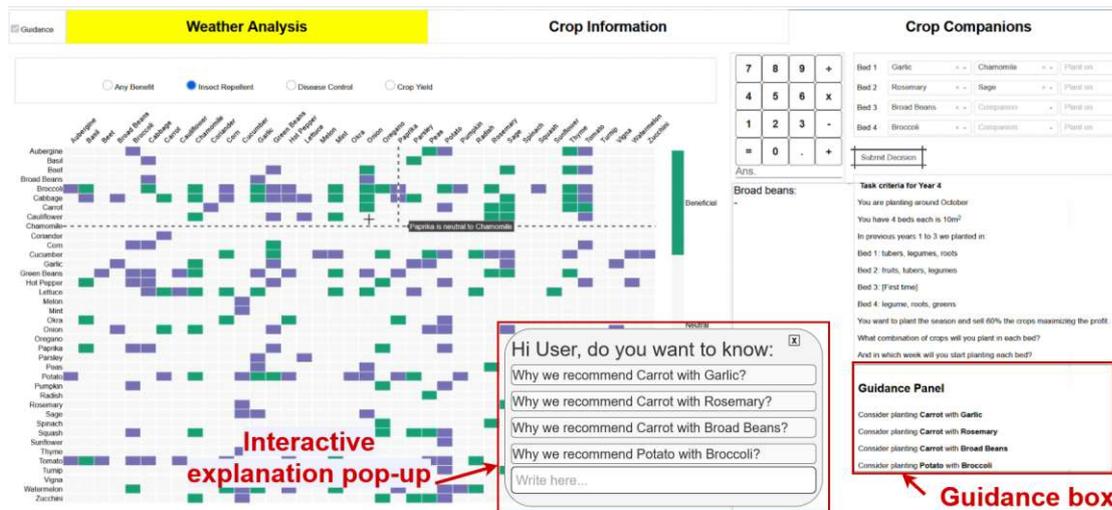


Figure 5.5: The resulting guidance explanation for the first scenario.

confidence. Thus, the ■ *purpose* of the explanation could be to either *improve decision making* or *boost user confidence* (Figure 5.4, *Why*). Nevertheless, the provenance data indicate that although the user's decisions might not be wrong, the user's confidence seems low. Thus, *boosting user confidence* is more appropriate, as the user may still hold on to their current plausible decisions. Finally, the designer identifies ■ the *guidance goal* as the *knowledge generation*, for example, by prescribing optimal decisions to the user. Moreover, the ■ *guidance degree* and ■ *task type* are identified as prescribing guidance within a domain task [CGM⁺18], respectively (Figure 5.4, *Which*). In the case of competing context responses, for example, for the goal of the explanation, the designer or system must select one response based on the best judgment of trade-offs. The discussion below elaborates on how this context further impacts the design.

The designer uses the framework to construct the ■ *structure*, according to the model in Figure 5.3. At first, the designer reasons about the structure of the explanation in terms of its content and visual appearance (red box in Figure 5.3). Since the guidance supports a *domain task*, an *empirically adequate* or *justified* explanation is required, i.e., the ■ *justification* (Figure 5.4, *What*). This design decision can also *boost users' confidence* in their decision, i.e., the *explanation purpose* (Figure 5.4 *Why*). Given the nature of empirical knowledge in the eco-farming domain, an *empirically justified* explanation is more applicable. The designer characterizes the ■ *explanandum type* as the input of the guidance to support better the *accuracy* of the explanation. Moreover, an ■ *explication* that illustrates the impact that ■ an *input* of one sub-decision has on other subsequent sub-decisions is essential to support the *transparency* of the explanation, allowing the user to understand how their input affects guidance and impacts their future decisions.

A higher degree of explanation is essential to give users confidence in their decisions and support them in their upcoming decisions. The dashboard provenance is continuously recorded and analyzed to identify a fall in trust when it happens. The system presents

the explanation when a trust ■ *gap is identified* through the provenance data during the analysis (Figure 5.4, *When*). This helps users ground their decisions in accepted knowledge of eco-farming and boosts their confidence. Also, given that the user is a *novice* user, an explanation of the algorithm might confuse them and lead to a further lack of confidence. It might also complicate the explanation by diverting from the required *simplicity*.

To balance *simplicity* with *accuracy* and *transparency* in the explication, the designer decides to implement ■ *an interactive dialogue mode* that can allow the user to navigate the elaborate justification within their capacity without overwhelming the user. The explication utilizes the *verbal* ■ *medium* without employing the visual channels to avoid cluttering a visually dense dashboard. This maintains the *simplicity* of the explanation (Figure 5.4, *How*). Moreover, since the system explains the *input*, rather than the algorithm, explanations do not need to be *numerical*. A ■ *verbal* explanation is more practical, especially in an *interactive dialogue*.

In the last step, the designer must decide the ■ *setting* of the explanation (Figure 5.3)—specifically, where to present the explanation (Figure 5.4, *Where*). Since the explanation will be presented as a *verbal interactive dialogue* with the user (Figure 5.4, *How*), the designer decides that it is appropriate to present the explanation as a ■ *pop-up*. A *separate box* is a possibility. However, to balance the high degree of explanation with its *simplicity*, the *non-intrusive pop-up* would allow the user to move the explanation box freely as needed while gradually seeking an explanation and building a clear understanding of the guidance provided (see pop-up at the bottom of the dashboard, Figure 5.5).

5.7.2 Use-Case Scenario 2: Guidance in Therapy Planning

The second case is based on the evaluation of a guidance-enhanced proton therapy planning system in chapter 3. The VA system enables users to decide between several alternative plans for treating cancer patients. The evaluation highlighted the need to add explanations that maintain and boost the users' trust in the system's suggestions, such as when the system proposes a slice of interest to explore. The system presents guidance on the slice level as a grey-scale heatmap (see Figure 5.7). This ■ *guidance orients* the user towards the interesting 2D slices to compare within the patient's 3D volumetric data. During an evaluation interview, one user—a ■ *knowledgeable researcher*—did not find this guidance helpful. Ideally, the designer should have added an icon to allow an explanation of the guidance ■ *on demand*. Taking the role of designers, we demonstrate how we would have designed adequate guidance explanations based on our model.

The designer starts by identifying the ■ *dynamics of user trust* (Figure 5.6). The guidance is provided in the ■ *early stage* of the analysis, as the user has not seen the VA system or guidance before. At this point, the user mainly expects the guidance to be ■ *usable*, i.e., the guidance will allow the rapid detection of slices of interest for the analysis. Thus, the designer decides to prioritize the ■ *interpretable* characteristic of the explanation (Figure 5.6, *Characteristics*). To this end, explanations have to be *anthropomorphic*,

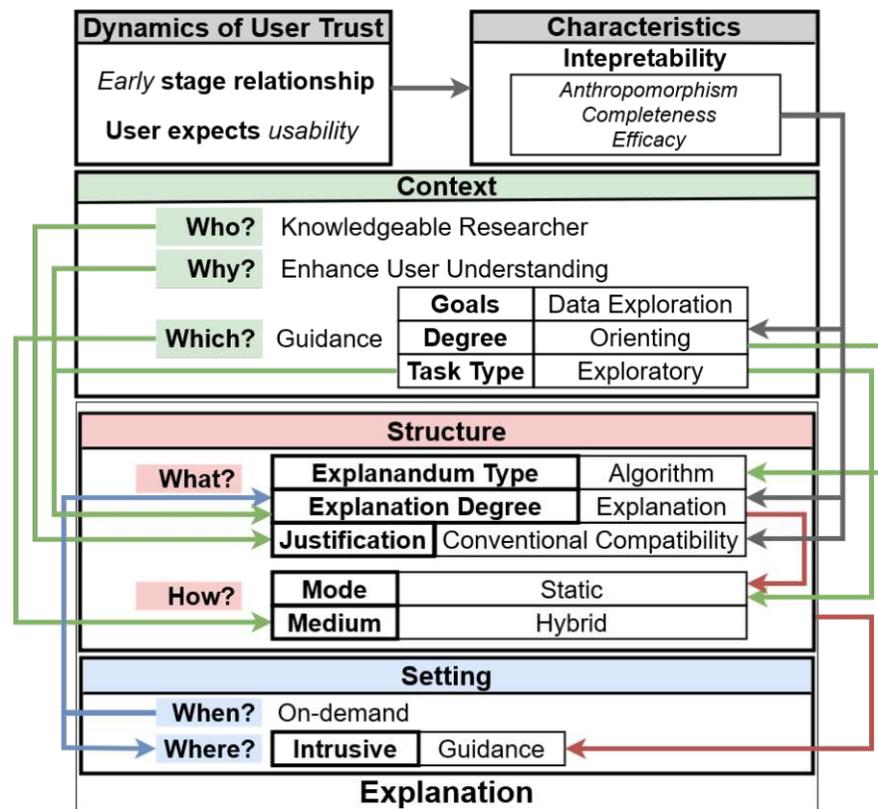


Figure 5.6: The flow of reasoning of a designer for the second use-case scenario: guidance in a therapy plan VA system [MMT⁺23] following our model.

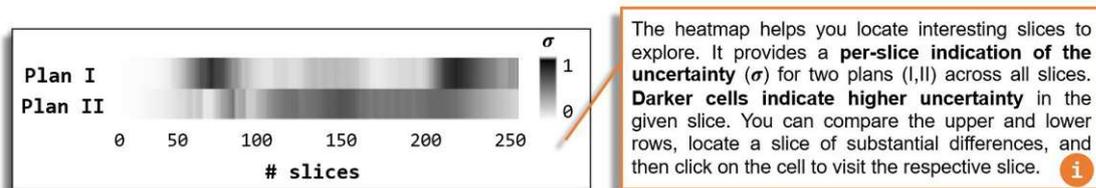


Figure 5.7: The resulting guidance explanation for the second scenario [MMT⁺23].

complete, and yield *efficacy* (see also Figure 5.3). Then, the designer characterizes the ■ *context* of the explanation. As mentioned above, the user is a ■ *knowledgeable researcher* (Figure 5.6, *Who*) who needs to understand a guidance of ■ an *orienting* degree that supports the goal of *data exploration* while performing an *exploratory* task (Figure 5.6, *Which*). The explanation aims to ■ *enhance understanding*, as the user declared that they could not understand the guidance (Figure 5.6, *Why*).

Next, the designer determines the ■ *structure*, i.e., *what* explanation is needed and *how* it should be presented (red box in Figure 5.6), along with the ■ *setting*, i.e., the *when* and *where* aspects of an explanation (blue box in Figure 5.6). The *orienting* guidance aims to support the *exploration* of the alternative treatment plans. Specifically, the user

needs to understand the *algorithm* used to compute guidance (■ *explanandum type*). This choice will enable the user to understand the guidance encoding as intended, i.e., contributing to its *efficacy*. Since the user is *knowledgeable researcher*, the designer decides that the explanation has to be *compatible with the conventions* in the medical domain (■ *justification* in Figure 5.6) to enhance its *efficacy*. Given the data and model’s uncertainty, a higher level of justification is not possible. The system should provide either ■ an *explanation* or an *explication* to ensure the *completeness* of the explanation in light of the *user demand* for it. In this scenario, an *explanation* is sufficient, given that 1) the guidance aims to *enhance a user understanding*—not to *promote insight discovery*—and 2) the guidance supports an *exploratory task*—not a *domain task*. Given the specific problem, the designer determines that an explanation should be presented *when* requested. Thus, the system could provide an ■ *intrusive* explanation as a tooltip *where* it was demanded (Figure 5.6, *When* and *Where*).

The delivery manner of the explanation should also be determined (Figure 5.6, *How*). The guidance described in this scenario comprises a heatmap of an uncertainty indicator for each slice [MMT⁺23]. Thus, a ■ *hybrid of verbal and numerical medium* of explanation is needed. The numerical component can be used to explain the formula of the uncertainty indicator encoded in the heatmap cells, and the verbal components support an understanding of how the guidance is read and used. For the ■ *mode* of the explanation, the designer concludes that an appropriate mode to present the explanation is *static*. A *dynamic* or *interactive dialogue* mode would have been more appropriate if the guidance supported a domain task or if an explication was required to accommodate longer (or more complex) explanations. However, a dilemma arises here: earlier, the designer determined that the explanation would be more effective as *anthropomorphic*. The designer settles the dilemma by deciding on a *static* explanation since the latter is more appropriate for the given *explanation degree* and *guidance task type*. Still, the *anthropomorphic* quality can be pertained by delivering the *verbal* and *numerical* explanation using a second-person point of view and personification [QZC22]. Based on the structure above, the explanation is best delivered as a tooltip ■ *on the guidance* (Figure 5.6, *Where*) when the explanation is *demanded* (Figure 5.6, *When*). The explanation would appear as shown in Figure 5.7.

Use-Case Scenario 3: Guidance in Time-Series

The third case is based on a guidance-enhanced approach by Ceneda et al. [CGM⁺18]. The work evaluates a guidance approach to support pattern detection within a spiral plot (see Figure 5.9). The guidance offers ■ *orienting* support to detect cyclical patterns in time-series data. The resulting guidance appears on the control slider as a delineation of areas of interest, as well as precise indications of cycles that the user could investigate. The evaluation in the paper highlighted that users need to receive additional support to understand how the guidance was calculated. Taking the role of designers, we show how we would have designed the explanation a posteriori.

We adopt our context-aware model from Figure 5.3. Figure 5.8 depicts the resulting

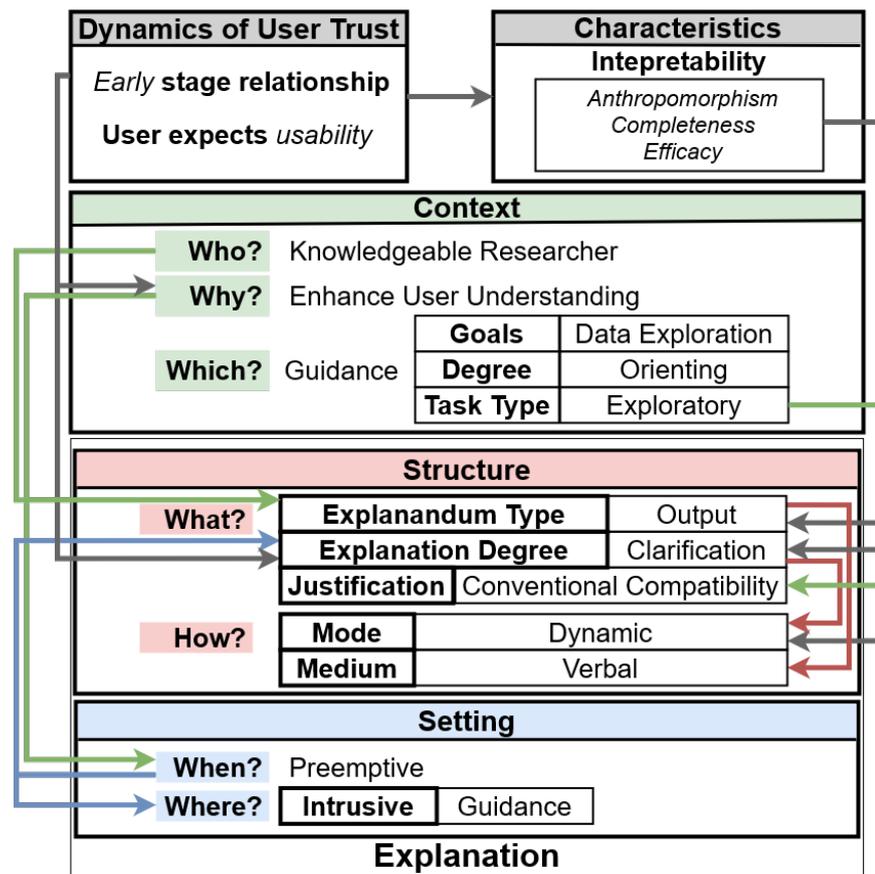


Figure 5.8: The reasoning flow of a designer for the first use-case scenario. The designer designs explanations for guidance in time-series representations [CGM⁺18] following our context-aware model (Figure 5.3). Decisions might be interlinked, as indicated by the arrows.

explanation design. The designer starts the work by identifying the *stage of the user's trust* based on the domain and application. In this case, the stages can be set based on the insights retrieved or the temporal length of the analysis if the former cannot be measured or externalized. Subsequently, the designer has to identify possible issues concerning the explainability of guidance in different analysis stages. Let us assume the user is in their ■ *early stage* of the analysis. During the early stage, the user might have several ■ *expectations* (see Figure 5.3); among which the designer prioritizes *usability* (Figure 5.8, *Dynamics of User Trust*). Although this expectation might change in later stages of the analysis, for example, later, the user may rather expect consistency later, and the explanation algorithm should be able to adapt to this change of expectations, we do not consider this case further within this use-case scenario for the sake of simplicity. Based on Figure 5.3, when prioritizing *usability* at an *early stage* of the analysis, the designer of the XG should strive for ■ *interpretability*. It means that the system should deliver an explanation that should be characterized by *efficacy*, *completeness*, and *anthropomorphism*

5. TRUSTME: A CONTEXT-AWARE EXPLAINABILITY MODEL TO PROMOTE USER TRUST IN GUIDANCE

(Figure 5.8, *Characteristics*). We explain how the designer can pursue these characteristics below.

The designer moves to identify the ■ *context* in which explanations will be provided. As mentioned in the paper of Ceneda et al. [CGM⁺18], the users of the guided system are PhD students familiar with the spiral plot, but they are not domain experts. Thus, the designer frames them as ■ *knowledgeable researchers* (Figure 5.8, *Who*). Furthermore, the evaluation participants reported an issue with understanding the motivation for the guidance. Thus, the ■ *purpose* of the explanation could be to either *facilitate user analysis* or *enhance user understanding* (Figure 5.8, *Why*). However, being in the early stage of the analysis, *enhancing user understanding* is more appropriate, as facilitating user analysis requires some familiarization. Finally, the designer identifies the ■ *guidance goal* as the *facilitation of data exploration*, for example, by pointing the users to areas of interest, and the ■ *guidance degree* and ■ *task type* as orienting guidance within an exploratory task [CGM⁺18], respectively (Figure 5.8, *Which*). In the case of competing context responses, for example, for the goal of the explanation, the designer or system must select one response based on the best judgment of trade-offs. In the discussion below, we elaborate on how this context further impacts the design.

Subsequently, the designer uses the framework to construct the ■ *structure*, according to the model in Figure 5.3. At first, the designer reasons about the structure of the explanation in terms of its content and visual appearance (red box in Figure 5.3). The designer characterizes the ■ *explanandum type* as the output of the guidance (since users have problems understanding it) by using a conventionally compatible clarification, i.e., the ■ *justification* and ■ *explanation degree*, respectively (Figure 5.8, *What*). A higher degree of explanation might overwhelm the user engaging with the guidance for the first time. In addition, as understanding the guidance has been reported by the users to be a common issue, the designer decides for a ■ *preemptive* explanation (blue box in Figure 5.3) that is provided to the user as soon as they first encounter the guidance (Figure 5.8, *When*). Since *completeness* was identified above as one of the necessary explanation characteristics, a *clarification* is chosen as the best course of action instead of a *tacit* explanation. Also, given that the user is a *knowledgeable researcher*, an explanation of the algorithm is unnecessary. The *efficacy* of the explanation can be achieved by explaining the *output* of the guidance. Finally, a *conventional compatibility* is enough since the guidance supports an *exploratory task*, not a *domain task*. Furthermore, the designer decides to present the clarification in a *dynamic* ■ *mode* and *verbal* ■ *medium* without employing visual channels (Figure 5.8, *How*). Again, several aspects of the model are interlinked (arrows in Figure 5.8). For instance, *referencing* has been chosen to make the explanation *complete* by showing which component is related to the explanation. An interactive dialogue contributes to its *anthropomorphism* and, thus, to emphasizing its *interpretability*. Moreover, since the system explains the *output*, rather than the algorithm, a *verbal* clarification is more practical than a *numerical* explanation. A visual explanation should be rather avoided to avoid clutter.

Finally, the designer must decide the ■ *setting* of the explanation (Figure 5.3)—specifically,

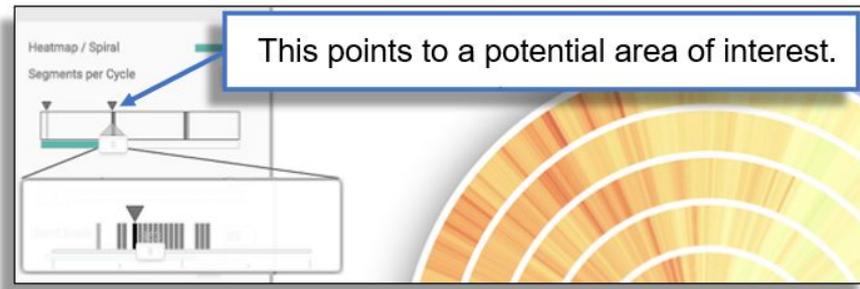


Figure 5.9: The resulting guidance explanation for the third scenario [CGM⁺18].

where to present the explanation (Figure 5.8, *Where*). Since the designer already decided that a *preemptive* explanation will be needed to help the user interpret the guidance as soon as it is provided to the user (Figure 5.8, *When*), the designer decides that it is appropriate to present the explanation *on the guidance* itself. A tooltip might not be sufficient, as an issue might arise before the user hovers over the guidance component. An appropriately preemptive encoding might be to position the explanation next to each guidance suggestion as soon as it appears. For example, above the guidance, an explanation could report: “*This points to a potential area of interest*” (Figure 5.9, blue box).

5.7.3 Design Walk-Through: Guided VA for Image Selection

To further demonstrate the applicability and usefulness of our model, we show how it can be applied in practice in a design walk-through. In this walk-through, an independent guidance designer assesses and redesigns his previously developed guidance-enhanced VA approach to include explanations. The expert has over 12 years of scientific, artistic, and professional research experience in visualization and VA, of which four years are in guidance. During this process, we acted as companions of the domain expert, presenting questions and alternative design choices. The actual end decisions on the design and the rationale behind them were always made by the expert. In the following, we describe the entire design process, reporting our step-by-step guidelines in the form of instructions, i.e., **the step-by-step objectives** that we presented to the designer, followed by their corresponding actions or reflections.

The considered VA system allows the user to detect unexploded ordnance (UXO) by selecting a set of images on the map, which are used to reconstruct (in subsequent steps) a specific area of interest [PMCM24]. In this endeavor, the user is supported with two guidance mechanisms, namely 1) one that orients the user by presenting interesting images for user consideration, and 2) one that prescribes a given image set to the user, i.e., a full solution, to enhance the selection and ease the analytical process; see Figure 5.10. The designer already evaluated the original guidance with six practitioners with varying expertise levels, as shown in the paper [PMCM24, p. 72]. The evaluation highlighted that participants with more expertise showed “*agonistic behavior*” towards the prescribing

5. TRUSTME: A CONTEXT-AWARE EXPLAINABILITY MODEL TO PROMOTE USER TRUST IN GUIDANCE

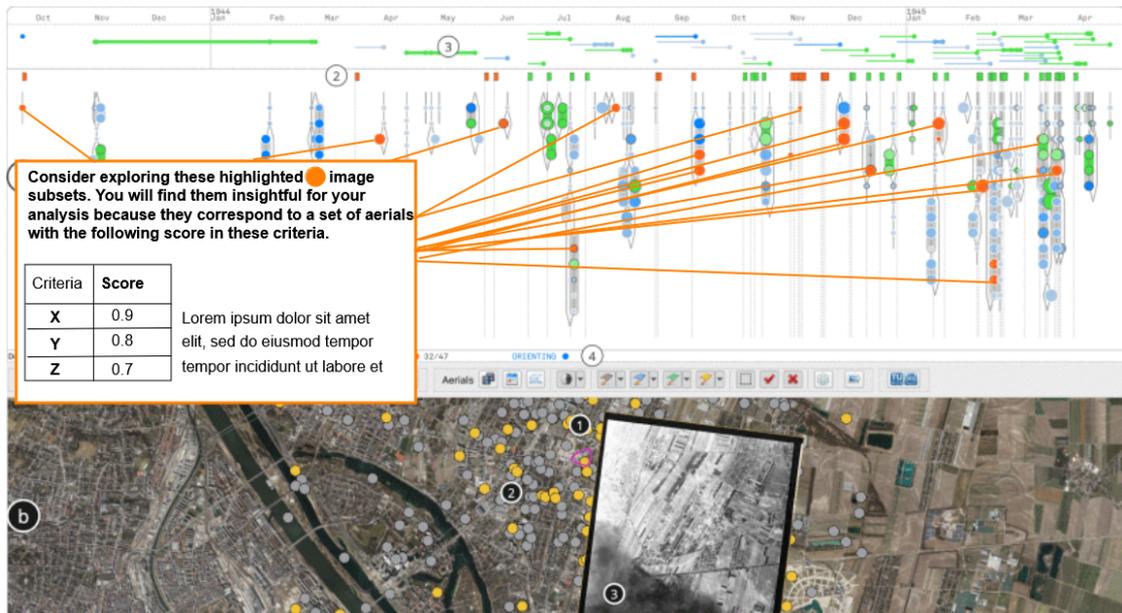


Figure 5.10: UXO guided-VA interface: the oriented guidance is encoded in the blue dots and the prescribing guidance in the orange dots [PMCM24]. The explainability in the orange box with the connecting orange lines is the product of the design session conducted a posteriori with the expert designer.

guidance and a need to redesign the system to include appropriate explanations to accompany the guidance.

■ *STEP 1: Context examination.*

Objective: “Outline the main goal of guidance, guidance degree, and task type supported by the guidance (**Which?**). From the evaluation, describe the target user(s) based on the five user categories in the green box in Figure 5.3 (**Who?**). Also, describe why the explanation is needed for each user type (**Why?**). This step is important to define the context. If an explanation is needed for different users, connect the reason for the explanation explicitly to the respective user type.”

The guidance with explainability issue is the *prescribing* guidance. It manifests as orange dots in the interface to recommend an image subset to the user. This aims to promote *knowledge generation* and discovery in a *domain* task, as seen in Figure 5.10. As stated by the designer, the issue occurs during the expert practitioners’ evaluation. The explanation is needed to *enhance the user’s understanding* of the guidance.

■ **STEP 2: Characteristics selection.**

Objective: “Define the three stages that describe the trust relationship phases between the user and the system (early, mid, and late stages). Based on the evaluation and context examined in Step 1, describe the users’ expectations from the guidance tool and at what stage of the trust relations these appear. Use the top-left matrix in Figure 5.3 to select the primary characteristic of guidance. From the tables in the bottom left box, retrieve the secondary-level characteristics. These characteristics guide the design decisions. Refine the reason the explanation is needed (**Why?**).”

The designer is interested in improving the explainability of the guidance in the *early stage* of the trust relationship since it is difficult to regain an expert’s trust if it is lost. An expert practitioner expects *competency* from the guidance, and as such, the primary explanation characteristic that he will focus on is *interpretability*. This characteristic is boosted by improving three secondary characteristics: anthropomorphism, completeness, and efficacy.

■ **STEP 3: Structure description.**

Objective: “Implement the analysis comprehensively, reasoning about the appropriate explanation content (**What?**) and form (**How?**), as indicated in the red box in Figure 5.3. This should account for the context anatomy from Step 1. Generally, design the first draft of the explanation strategy based on the context information. Use the information from the explanation’s content (**What?**) to define the form (**How?**) while considering the anatomy of the guidance (**Which?**).”

The designer decides that explaining the *output* is more important to the expert user. Given that the explanation should enhance the understanding, he thinks providing an *explanation* is better. Given the guidance context, an explication is not necessary. Moreover, an empirically adequate explanation is not possible; thus, he decides on the second most evident justification, *empirically justified*. With an empirically justified explanation, he believes that a *hybrid* explanation, *verbal* and *numerical*, is important. The explanation is better shown in a *static* form.

■ **STEP 4: Setting definition.**

Objective: “Describe when and where the explanation will be presented to the user (bottom box in Figure 5.3). The temporal setting is influenced generally by the stage of trust defined in Step 2 and the reason the explanation is provided (**Why?**). Define the spatial setting of the explanation based on the temporal setting and structure described in Step 3.”

The designer does not want to lose the expert’s trust early if they judge the system prematurely. Thus, he wants to explain *preemptively* as soon as the guidance is presented. With this purpose in mind, the explanation is more effective when provided *intrusively* on the *guidance* visual space.

■ **STEP 5: Structure refinement.**

Objective: “Revise the content of the explanation considering the temporal setting defined in Step 4 and trust dynamics defined in Step 2. Then, refine content and form design to exemplify each of the three secondary-level characteristics.”

Since the trust issue occurs in an early stage, a preemptive explanation is important. Thus, the designer also concurs that the explanation is better presented as a hybrid, adding a visual referencing that can help the user understand the explanation’s reference. This approach can support the efficacy of the explanation. The verbal justification is re-drafted using a second-person point of view, addressing the user, to improve its anthropomorphism characteristic. Initially conflicted between providing an empirically justified or conventionally compatible explanation, the designer is now convinced that an empirically justified explanation is needed to support the completeness of the explanation. The designer sketches the possible explanation, as illustrated in Figure 5.10.

5.8 Discussion

Differences between XG and XAI When building the framework, we often reasoned about the commonalities between XAI and explainable guidance (or even drew upon literature from the former for the latter). Although explainable guidance intertwines with XAI in several notions and sometimes in their application, designing explanations for guidance mandates distinct approaches and investigates questions different from XAI. Thus, we should consider unique aspects, such as user need for guidance, in constructing explanations—especially when the final goal is to boost trust in guidance. For example, although a guidance-enhanced VA system may *use AI* to generate guidance, this is only occasionally the case [CGM19b]. Considering this factor and the differences discussed previously, the gap between XAI and XG can be bridged by systematically questioning the context and goals of the explanation.

Mixed-initiative explainability Another difference between XG and XAI is that guidance is a mixed-initiative process. In such a case, the system should consider both interpretability and explainability—from the user’s and system’s perspectives. Guidance alternates in a loop between the user and the system to complete the analysis, and both interpretability and explainability notions are required [HR19]. As we clarified in section 5.3, explainable systems aim to convey a precise interpretation of the visual

output, while interpretable systems focus on conveying one (or more) plausible meaning(s). Although the notion of interpretable guidance contradicts the accuracy and precision requirements for effective guidance, the notion of multiple plausible interpretations of guidance can be an exciting topic. Studying whether guidance can fulfill multiple aims can forge paths for more agile and efficient methods to address design issues for various user contexts in guidance.

Reproducibility The expert design walk-through highlighted an expectation of reproducibility. Reflecting on the process described in subsection 5.7.3, the expert believed that the design process was useful for pinpointing the issue and understanding the target of the explanation. However, he also suggested that the framework should be accompanied by term definitions to make it effective. During the session, the designer was stressed about correctly answering the questions in each step. However, the model maintains a space for the designer’s creativity. There is more than one alternative when designing XG. Ideally, designers should reason while making design decisions. Then, the designed explainability, or several, should be evaluated for efficacy and efficiency. The literature proposes approaches to evaluate guidance [CCEA⁺24], which can be extended for evaluating explanations with calibrated metrics and targets. Yet, a challenge could be the lack of approaches to accurately and objectively measure subjective metrics, such as user confidence [MCER25].

The evaluation of guidance also highlights issues with regard to users’ expertise level, as experienced in the design walk-through in subsection 5.7.3. We argue that designing an explanation should account for diverse users. This is not different from user-specific considerations in designing VA or the respective guidance. In this paper, we highlight the importance of designing context-aware explainability strategies. This approach demands that the designers make trade-off decisions between several aspects of the explainability strategy. The distinction of characteristics introduced in the model aims to facilitate those decisions. For example, the conflict between mitigating the cognitive load and completeness of the explanation is resolved when the matrix helps the designer decide on the critical characteristics for the design context (Figure 5.3, top grey boxes). Finally, XG strategies are not only about designing an explanation but also about deciding whether you need the explanation. If the extra information does not add to the efficiency or effectiveness of the tool, then it should not be included to avoid cognitive overload. Thus, the designer should evaluate the guidance with target users to learn issues in trusting or using the guidance.

Moreover, our proposed XG model is designed to address the practical questions of guided-VA designers and the acceptance issues among users. To address the former, we need to balance between providing a detailed model to improve the practicality of the model and reducing unnecessary details to preserve a creative space for the designer. We also reduce unnecessary details to avoid prejudices about users’ needs and behavior. The model addresses the effectiveness of user analytical behavior by pivoting the line of design decisions to user expectations. Thus, the model addresses the gap between

5. TRUSTME: A CONTEXT-AWARE EXPLAINABILITY MODEL TO PROMOTE USER TRUST IN GUIDANCE

user expectations and system behavior, forging a more coherent guided-VA user–system ecosystem.

Limitations Although explanations are seen as an improvement to the analytical visualization, the literature points to several *possible biases and pitfalls* [EAM22]. Some of the common biases are *knowledge bias*, i.e., the presupposition of false assumptions on user’s knowledge background; *attribution bias*, as in defining user groups wrongly; *availability bias*, which concerns the availability of explanation; *confirmation bias* by the user who looks for explanations that agree with their views; *explainability of explanation*, which refers to unclear explanations that cannot be adequately interpreted; and *inaccurate explanation* that could mislead the user in the analytical process, illustrated in Figure 5.11. Additionally, we recognize that providing a layer of explanation to the guidance could result in a *heavy reliance* on the explanation. This reliance may strip the user of the agency in the analytical process or defy the aim of the guidance, which is to support and facilitate the user’s analytical process, not replace it [CGM⁺17b]. Finally, the characteristics and aims we propose in our model are a starting point. The literature on explainability presents several characteristics and aims, which could *expand our model*. This model should also be calibrated through accompanying rigorous studies. Future work should evaluate how implementing this model impacts user trust while enhancing the overall experience in comprehensive guided VA scenarios.

Accountability and transparency are two ethical challenges posed in the process of achieving explainability and interpretability. In high-risk VA applications, users might be more open to receiving elaborate explanations to mitigate potential safety risks, such as in medical applications and autonomous driving [Div24]. Explainable models are more relevant in these contexts; however, the high-risk context also poses more restrictions

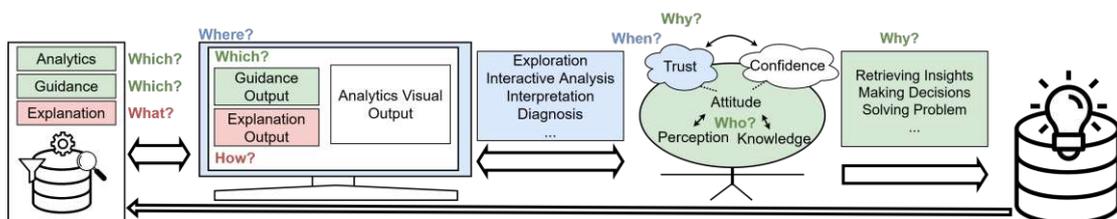


Figure 5.11: A workflow of explainable guidance in VA illustrating the interplay of explanations within the analytical process. This workflow illustrates the different contexts and how this information influences the structure of explanations. The data is filtered, processed, and analysed in the back end. Guidance and its explanation are generated during the interactive VA process—when needed—and are visualized on the screen. The user may influence the guidance and explanations using interaction. Explanations affect the user’s trust, which, in turn, impacts the user’s perception and knowledge. Eventually, the user fulfills the main task of the process, which is to produce new knowledge and data that feed into the system in the next VA task.

on the design of explainability models. For example, the use of generative AI in certain scenarios might pose a liability issue.

The proposed model attempts to serve as a comprehensive explainability model. This pursuit comes with trade-offs. A complete model might overwhelm the designer. It can also be challenging to comprehend. However, to mitigate the potential drawbacks of the model completeness, we explain all concepts presented in the model, as seen in Table 5.1. The exemplified use case scenarios also attempt to provide an onboarding tool to guide designers. It might not be possible to address all trade-offs of a comprehensive model fully. The model is presented in several color-codified parts to allow future researchers to adapt and further refine this model. However, this remains a limitation of the model that should be addressed in future studies. Further evaluations could show whether, why, or how such a model could be simplified to specific domain contexts. Potentially, designers can adopt different subsets of this model. Conducting a future study targeting guidance designers can shed more light on the usability of the model and how it can be improved. As this is the first attempt at an XG model, we expect an active refining process by experts in the field.

Summary This chapter proposes a framework for explainable guidance and discusses the factors that contribute to its design. Future research should look into methods to evaluate XG, accounting for specific purposes and contexts. To this end, methods to externalize user trust in real time are essential. Moreover, a helpful prospect is constructing a model that can integrate explainability models for both XAI and XG. Finally, considering mixed-initiative systems, integrating interpretability into the explainability model could refine the explanations further.

Conclusion

This work aims to enable VA researchers to address uncertainties in VA decision making, namely subjective uncertainties, by applying effective guidance approaches in a way that benefits their trust and confidence in the analytical process. Subjective uncertainties crucially impact how users make their decisions. In a data-driven world, VA is taking an increasingly bigger role in decision making in multiple domains by combining the computational power offered by modern computers, algorithms, and the reasoning of humans through interactive visual approaches. Thus, VA designers face the challenges of designing effective tools to address uncertainties in VA decision making and allow users to make optimal decisions. We explore guidance as an effective approach to address uncertainties in VA decision-making scenarios where decisions cannot be measured for correctness. We also analyze and address the critical interplay of user trust and confidence and their effect on those decisions.

6.1 Reflection

Here, we reflect on the research we conducted through our work to respond to the research questions outlined in section 1.2. We presented a two-dimensional uncertainty guidance to support the comparison and uncertainty analysis in PT planning visualization. We exemplify uncertainty guidance on a PT planning dashboard for plan comparison. Moreover, an evaluation measured confidence and insights from the use of uncertainty guidance. In our work, we conceptualize *user confidence* and distinguish it from *user trust*. We then proposed SNA metrics to evaluate user confidence in VA under uncertainty and conducted a study to examine the value of the proposed metrics. Finally, we propose a framework for XG and discuss the factors contributing to its design, such as user and domain contexts.

RQ1 How can we design effective guidance to boost experts' confidence in their decision making under the presence of uncertainty?

The evaluation of our proposed PT application in chapter 3 yielded encouraging initial results and provided insights for forthcoming directions. We detected a gap in evaluating guidance approaches in VA under uncertainty in the clinical domain, where conventions are firm and analytical processes well-established (and often inflexible). The effectiveness of guidance in such a scenario relies on boosting the expert's trust in the system. Moreover, to properly evaluate the effectiveness, we must be able to measure the levels of user confidence in their decision (and throughout the analytical process). Thus, we subsequently explored the evaluation of *user confidence* in chapter 4 to account for the domain-expertise factor. We also aim to continue developing our guidance mechanisms to reflect better the lessons learned from the current work. Our work constructs a guidance mechanism by understanding the user's needs and the detailed levels of the data explored. It attempts to guide the user through the uncertainties on multiple data resolutions as needed.

RQ2 How can we effectively evaluate the impact of guidance techniques on user confidence?

Chapter 4 explored the potential of provenance techniques in informing the evaluation of uncertainty guidance. We envision that our approach (with modifications) can be generalized to different uncertainty visualization applications, such as the medical domain, weather forecasting, financial prediction, or material sciences. The global characteristics, global topology, and local topology metrics we proposed in chapter 4 offer the potential for an insightful analysis of user confidence in VA. They can complement the commonly used self-reporting surveys. With further modeling and tuning, our proposed metrics could improve the assessment of user confidence and, consequently, support a better evaluation and development of VA tools, including guidance.

Knowledge-assisted visual analytics (KAVA) literature did not establish a clear distinction between the nuances of user confidence and user trust. Our work contributes to setting more precise boundaries to assist a better study of the interplay of trust and confidence in VA decision making. We promote the use of quantitative metrics to measure user confidence and overcome the limitation of subjective bias in self-reported results. However, our work also emphasizes incorporating qualitative evaluation to complement the quantitative approach. Although KAVA research relies increasingly on analyzing provenance, we argue for the value of network representation and SNA metrics. We view actions and components of VA as entities that engage in an exchange of knowledge that resembles a social exchange via user interactions.

RQ3 How can we improve users' trust in the guidance techniques?

We developed a framework for explainable guidance to study the impact of explainability on *user trust*, which impacts their confidence in sense making and decision making. The

framework was built with extensive research on explainability in different subfields of VA and AI. It adapts the requirements of explainability to research findings on guidance in the VA. The framework proposes a structured guidance that prioritizes cognitive relief, clarity, timeliness, and interpretability. We demonstrate its potential to support the trustworthiness of guidance in three literature scenarios and a design walk-through with a guidance design expert. While validated with use case scenarios and expert design walk-through, the practical applicability of the framework requires further studies. Namely, a guided-VA application designed using the framework should be evaluated for effectiveness by focusing on the user trust component.

While prior literature on KAVA focuses on building trust through transparency and constant feedback, our work prioritizes understanding the user and the dynamics of user trust in decision making. Our work proposes structured explanations that do not always need to be interactive, as prioritized in the field of VA explainability. Nonetheless, we incorporate the value of user feedback into an adaptive explainability model. This adaptive feature can be further strengthened by incorporating methods to externalize user trust.

6.2 Outlook

Uncertainties in VA The topic of guidance in VA has gained considerable attention in recent years. VA approaches are fast becoming crucial tools to enable human-AI collaboration, address subjective uncertainties, and optimize decision making in multiple domains. As our understanding of the VA decision-making process evolves, we should consider mapping the VA cycle to more than the traditional four stages: data, hypothesis, visualization, and insight [GMR⁺23]. It is also worth considering the different scenarios of decision making. For example, consider a scenario where the user of a VA system does not make the final decision; instead, they have to communicate insights to a third party and engage in a back-and-forth VA process to reach a decision in a collaborative effort. The direct user of the system also becomes a *proxy user* who brings input from another person to the analysis process, with all its attached biases. This hidden layer of VA usage can have a crucial impact on the VA decision-making process, the layers of uncertainty, and the type of guidance the direct user needs. Thus, revising the traditional VA cycle to include new stages can help rethink VA guidance.

Moreover, our understanding of subjective uncertainties grows as we understand the user better. Currently, the uncertainty taxonomies in the literature fail to provide a comprehensive picture of how uncertainties impact decision making in VA. Future research should focus on re-mapping uncertainty events and their impact on decision making. This research allows designers to understand better how to address these uncertainties. Specifically, a comprehensive study of the properties of the user attitude (also beyond trust and confidence) and their interplay in the VA decision making allows designers to propose more effective context-aware and uncertainty-aware guided-VA systems.

User attitude in VA An interesting direction for future work in evaluating guidance would be to explore the distinctions between confidence and trust—quantitatively and qualitatively—within the specific visual data analysis context to further understand their respective impacts on user interactions and decision-making processes. Studying the impact of XG on the user’s trust and confidence, especially in complex uncertainty scenarios, can be valuable. Moreover, future research should look into methods to evaluate XG, accounting for specific purposes and contexts. To this end, developing methods to externalize user trust in real time is essential. An accurate quantification of user attitude properties can allow us to explore whether and which levels of trust can promote an efficient and effective analytical process. Related to that, it allows us to investigate the appropriate trade-offs between effectiveness (making the user confident enough) and efficiency (allowing the user to complete the tasks in a timely manner). Moreover, an interesting path for investigation lies in understanding how changes in user attitude properties, namely trust and confidence, introduce uncertainties to the VA process and to map out these uncertainties. Also, it would be helpful to understand how these uncertainties impact and change the properties of user attitude in the VA cycle.

Furthermore, a helpful prospect is constructing a model that can integrate explainability models for both XAI and XG. Considering mixed-initiative systems, integrating methods to uphold interpretability into explainability models could further refine explanations’ effectiveness. The user’s understanding of guidance cues profoundly impacts user trust. Thus, the mechanism of user interpretation should be studied in connection with the research on user attitudes. VA research should involve more social psychology studies in a more integrated research ecosystem. It is not about bringing in lessons from the social psychology field; rather, it is about re-imagining the social psychology of today’s humans, who use visual systems to make data-driven decisions daily.

Agency in VA Moreover, in the age of rapid AI development, studies should carefully consider the optimal balance between human and machine agencies. Guiding the user through uncertainties in decision making should not strip the user of their agency. Furthermore, VA designers should not assume static interpretations of uncertainties in the VA process. Subjective uncertainties are dynamic and ever-changing [PFW⁺18]; thus, the guided VA should adapt to these changes while prioritizing the agent’s perspectives that carry the responsibility for the decision.

Understanding biases This dissertation discusses how user attitude impacts cognitive biases. However, it shies away from explicitly describing the mechanism by which these attitudes impact users’ cognitive bias, as it requires a more dedicated line of questioning. The dissertation distinguishes between two crucial properties of users’ attitudes in the study of visualization and VA: user confidence and user trust. This distinction sets the foundation for better describing how these properties of the user attitude impact cognitive biases. It is also worth exploring and describing what other attitudes impact

decision making in VA under uncertainty. Moreover, given that an attitude's properties can fall between cognitive and emotional biases, future work that clearly describes that interplay and intersection can provide a solid foundation to implement techniques that address users' attitudes.

From theory to practice Recently, the rise of Generative AI and large language models (LLMs) has opened the path for smart domain-agnostic guidance models that can generate context-aware guidance. However, this approach is still under-explored. Practical and ethical challenges remain to be addressed. On the practical side, we need to develop methods that can verify the trustworthiness of the output of these LLM models. Moreover, the effectiveness and efficiency of these models depend on high processing powers. Although such a problem can be addressed by cloud computing approaches, it still poses a challenge to VA decision-making systems in offline contexts. On the ethical side, one needs to consider the energy consumption and negative impact on the climate through the use of these models. Also, in fields such as the medical field, the input of patient data into third-party LLM models poses a privacy challenge. Thus, several questions remain to be addressed before generative AI can be widely used to generate guidance for VA decision-making.

Perspectives in VA One of the most significant limitations of the study of VA methods is that they are primarily centered in the global north. The perspectives of the global south are marginalized. Thus, future research in guided VA needs to incorporate more diverse perspectives to consider the nuances of user cultural and social contexts. This direction can create a more inclusive research environment, but most importantly, it can introduce new perspectives that can direct the research into exciting paths. For example, as we worked on our explainability model in chapter 5, knowledge produced in the global north led our investigation, methodology, and evaluation. An effective explanation's characteristics, context, and structure may differ from the perspective of the global south. On a higher level, in this dissertation, we relied heavily on work produced by researchers from the global north or who conduct their research in an institute in the global north. Almost all the papers included as references in this dissertation were published in platforms in the global north. The average number of authors per paper is 3.6. More than 70% of the first authors are from Europe or North America, and around 90% are from the global north. Around 89% of the first authors conduct their research in institutes in Europe or North America. More than 58% of the first authors work in institutes in one of three countries: USA (35.9%), Germany (11.7%), and Austria (11%). Around 70% of the first authors are men.

On the brighter side, we reference works of 275 unique first authors and 851 unique authors in total. The first authors spread across the six continents, conducting their research in 200 unique institutions located in 37 countries. In this dissertation, although 40% of the references were published in the past five years, we reference work that spans across the 20th and 21st centuries, dating back to 1901. More than 10% of the references were published before 2000. The work referenced is published in 102 different journals

6. CONCLUSION

and 48 different conferences. The Python code and worksheets used to generate these statistics are available in an OSF repository [Mus25]. These statistics of the 326 references in this dissertation highlight the importance of being conscious about our research's epistemological foundation despite time pressure. Researchers must also consider more diverse perspectives when evaluating their work, where possible.

Finally, this dissertation breaks out of the strict disciplinary boundaries and consciously incorporates diverse social and cultural perspectives from different research environments. The dissertation offers a broader understanding of guidance in VA. It positions guidance as a promising approach to address uncertainties, especially subjective uncertainties. Moreover, it establishes foundations to integrate the study of user attitude in VA processes and its impact on guidance, particularly the interplay of user confidence and user trust.

Bibliography

- [ACC⁺06] Vicky Arnold, Nicole Clark, Philip A. Collier, Stewart A. Leech, and Steve G. Sutton. The differential use and effect of knowledge-based system explanations in novice and expert judgment decisions. *MIS Quarterly*, 30(1):79–97, March 2006.
- [AGJ13] Herman Aguinis, Ryan K. Gottfredson, and Harry Joo. Best-practice recommendations for defining, identifying, and handling outliers. *Organizational Research Methods*, 16(2):270–301, January 2013.
- [AGN⁺20] Francesco Alzetta, Paolo Giorgini, Amro Najjar, Michael I. Schumacher, and Davide Calvaresi. In-time explainability in multi-agent systems: Challenges, opportunities, and roadmap. In *Explainable, Transparent Autonomous Agents and Multi-Agent Systems*, volume 12175 of *EXTRAA-MAS '20*, pages 39–53, Auckland, New Zealand, July 2020. Springer.
- [AKS21] Thomas Arnold, Daniel Kasenberg, and Matthias Scheutz. Explaining in time: Meeting interactive standards of explanation for robotic systems. *ACM Transactions on Human-Robot Interaction (THRI)*, 10(3):1–23, July 2021. Article no. 25.
- [AKvM⁺18] Lorenzo Amabili, Jiri Kosinka, Maarten A. J. van Meersbergen, Peter M. A. van Ooijen, Jos B. T. M. Roerdink, Pjotr Svetachov, and Lingyun Yu. Improving provenance data interaction for visual storytelling in medical imaging data exploration. In *Proceedings of the Eurographics/IEEE VGTC Conference on Visualization: Short Papers*, EuroVis '18, pages 43–47, Brno, Czech Republic, June 2018.
- [AR23] Saranya A. and Subhashini R. A systematic review of explainable artificial intelligence models and applications: Recent developments and future trends. *Decision Analytics Journal*, 7:100230, June 2023.
- [Arn06] David Arnott. Cognitive biases and decision support systems development: a design science approach. *Information Systems Journal*, 16(1):55–78, January 2006.

- [AS95] Christopher Ahlberg and Ben Shneiderman. Visual information seeking: Tight coupling of dynamic query filters with starfield displays. In *Readings in Human-Computer Interaction: Toward the Year 2000*, pages 450–456. Morgan Kaufmann, San Francisco, CA, second edition, 1995.
- [AZB⁺15] Syed Z. Arshad, Jianlong Zhou, Constant Bridon, Fang Chen, and Yang Wang. Investigating user confidence for uncertainty presentation in predictive decision making. In *Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction, OzCHI '15*, pages 352–360, Parkville, VIC, Australia, December 2015. ACM.
- [Bal01] James Mark Baldwin, editor. *Dictionary of Philosophy and Psychology*, volume 1. The MacMillan Company, New York, NY, 1901.
- [BAOL12] Ken Brodlie, Rodolfo Allendes Osorio, and Adriano Lopes. *Expanding the Frontiers of Visual Analytics and Visualization*, chapter A Review of Uncertainty in Data Visualization, pages 81–109. Springer, London, England, 2012.
- [Bar06] Terry Barrett. *Encyclopedia of Twentieth-century Photography: G-N ; Index. Vol. 2*, chapter Interpretation, pages 803–806. Taylor & Francis Group, New York, NY, 2006.
- [Bau19] Christoph Baumberger. Explicating objectual understanding: Taking degrees seriously. *Journal for General Philosophy of Science*, 50(3):367–388, September 2019.
- [BBC⁺20] Nadia Boukhelifa, Anastasia Bezerianos, Remco Chang, Christopher Collins, Steven Drucker, Alex Endert, Jessica Hullman, Chris North, and Michael Sedlmair. Challenges in evaluating interactive visual machine learning systems. *IEEE Computer Graphics and Applications*, 40(6):88–96, October 2020.
- [BBK15] J. S. Blumenthal-Barby and Heather Krieger. Cognitive biases and heuristics in medical decision making: A critical review using a systematic search strategy. *Medical Decision Making*, 35(4):539–557, May 2015.
- [BBR20] Stanislav Belyakov, Alexander Bozhenyuk, and Igor Rozenberg. Guidance in the visual analytics of cartographic images in the decision-making process. In *Advances in Soft Computing*, volume 12468 of *The 19th Mexican International Conference on Artificial Intelligence (MICAI '20)*, pages 351–369, New Mexico, Mexico, October 2020.
- [BBV22] Raissa Barcellos, Flavia Bernardini, and José Viterbo. Towards defining data interpretability in open data portals: Challenges and research opportunities. *Information Systems*, 106:101961, May 2022.

- [BDRD⁺20] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador Garcia, Sergio Gil-Lopez, Daniel Molina, Richard Benjamins, Raja Chatila, and Francisco Herrera. Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58:82–115, June 2020.
- [BFCA23] Dario Benvenuti, Matteo Filosa, Tiziana Catarci, and Marco Angelini. Modeling and assessing user interaction in big data visualization systems. In *Human-Computer Interaction*, volume 14143 of *INTERACT '03*, pages 86–109, Parkville, VIC, Australia, August 2023.
- [BGM19] Christian Bors, Theresia Gschwandtner, and Silvia Miksch. Capturing and visualizing provenance from data wrangling. *IEEE Computer Graphics and Applications*, 39(6):61–75, Nov.-Dec. 2019.
- [BH94] Henri Barki and Jon Hartwick. Measuring user participation, user involvement, and user attitude. *MIS Quarterly*, 18(1):59–82, March 1994.
- [BH19] Leilani Battle and Jeffrey Heer. Characterizing exploratory visual analysis: A literature review and evaluation of analytic provenance in tableau. *Computer Graphics Forum*, 38(3):145–159, June 2019.
- [BHJ⁺14] Georges-Pierre Bonneau, Hans-Christian Hege, Chris R. Johnson, Manuel M. Oliveira, Kristin Potter, Penny Rheingans, and Thomas Schultz. *Scientific Visualization: Uncertainty, Multifield, Biomedical, and Scalable Visualization*, chapter Overview and State-of-the-Art of Uncertainty Visualization, pages 3–27. Mathematics and Visualization (MATHVISUAL). Springer, London, England, September 2014.
- [BKD⁺23] Sarah E. Bergholtz, Sophia R. Kurnot, Melissa DeJonckheere, Sarah T. Hawley, Scott R. Owens, and Cathryn J. Lapedis. A mixed-methods study of clinicians’ attitudes toward pathology explanation clinics. *American Journal of Clinical Pathology*, 159(5):437–447, May 2023.
- [BKO⁺17] Zoya Bylinskii, Nam Wook Kim, Peter O’Donovan, Sami Alsheikh, Spanandan Madan, Hanspeter Pfister, Fredo Durand, Bryan Russell, and Aaron Hertzmann. Learning visual importance for graphic designs and data visualizations. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, UIST ’17, pages 57–69, Québec City, QC, Canada, October 2017.
- [BKR⁺17] Tanja Blascheck, Kuno Kurzhals, Michael Raschke, Michael Burch, Daniel Weiskopf, and Thomas Ertl. Visualization of eye tracking data: A taxonomy and survey. *Computer Graphics Forum*, 36(8):260–284, December 2017.

- [BLB21] Nadia Boukhelifa, Evelyne Lutton, and Anastasia Bezerianos. A case study of using analytic provenance to reconstruct user trust in a guided visual analytics system. In *IEEE Workshop on TRust and EXpertise in Visual Analytics*, TREX '21, pages 45–51, New Orleans, LA, October 2021.
- [BM99] Philip J. Barnard and Jon May. Representing cognitive activity in complex tasks. *Human-Computer Interaction*, 14(1-2):93–158, June 1999.
- [BM13] Matthew Brehmer and Tamara Munzner. A multi-level typology of abstract visualization tasks. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2376–2385, December 2013.
- [BMA⁺25] Camelia D. Brumar, Sam Molnar, Gabriel Appleby, Kristi Potter, and Remco Chang. A typology of decision-making tasks for visualization. *IEEE Transactions on Visualization and Computer Graphics*, 31(10):8536–8551, October 2025.
- [BMMC23] Iligim Dara Benoit, Elizabeth G. Miller, Ann M. Mirabito, and Jesse R. Catlin. Medical decision-making with tables and graphs: The role of cognition, emotions, and analytic thinking. *Health Marketing Quarterly*, 40(1):59–81, Jan.-Mar. 2023.
- [Bol08] Richard J. Boland. *Handbook on Decision Support Systems 1: Basic Themes*, chapter Decision Making and Sensemaking, pages 55–63. International Handbooks on Information Systems (INFOSYS). Springer, Heidelberg, Germany, 2008.
- [Bou08] Rabah Bousbaci. "Models of Man" in design thinking: The "Bounded Rationality" episode. *Design Issues*, 24(4):38–52, Autumn 2008.
- [Bro96] John Brooke. SUS: A quick and dirty usability scale. *Usability Evaluation in Industry*, 189(194):4–7, November 1996.
- [BVV⁺19] Tanja Blascheck, Lindsay MacDonald Vermeulen, Jo Vermeulen, Charles Perin, Wesley Willett, Thomas Ertl, and Sheelagh Carpendale. Exploration strategies for discovery of interactivity in visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 25(2):1407–1420, February 2019.
- [BWD⁺19] Christian Bors, John Wenskovitch, Michelle Dowling, Simon Attfield, Leilani Battle, Alex Endert, Olga Kulyk, and Robert S. Laramee. A provenance task abstraction framework. *IEEE Computer Graphics and Applications*, 39(6):46–60, Nov.-Dec. 2019.
- [CAA⁺20] Davide Ceneda, Natalia Andrienko, Gennady Andrienko, Theresia Gschwandtner, Silvia Miksch, Nikolaus Piccolotto, Tobias Schreck, Marc Streit, Josef Suschnigg, and Christian Tominski. Guide me in analysis: A framework for guidance designers. *Computer Graphics Forum*, 39(6):269–288, September 2020.

- [CAGM22] Davide Ceneda, Alessio Arleo, Theresia Gschwandtner, and Silvia Miksch. Show me your face: Towards an automated method to provide timely guidance in visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 28(12):4570–4581, December 2022.
- [Cam24] Cambridge University Press. Guidance | english meaning. In *Cambridge Academic Content Dictionary*. Cambridge University Press, 2024.
- [Cas04] Cristiano Castelfranchi. Trust mediation in knowledge management and sharing. In *Trust Management (iTrust '04)*, volume 2995 of *Lecture Notes in Computer Science (LNCS)*, pages 304–318, Oxford, England, 2004.
- [CAS⁺18] Christopher Collins, Natalia Andrienko, Tobias Schreck, Jing Yang, Jaegul Choo, Ulrich Engelke, Amit Jena, and Tim Dwyer. Guidance in the human-machine analytics process. *Visual Informatics*, 2(3):166–180, September 2018.
- [CCEA⁺24] Davide Ceneda, Christopher Collins, Mennatallah El-Assady, Silvia Miksch, Christian Tominski, and Alessio Arleo. A heuristic approach for dual expert/end-user evaluation of guidance in visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 30(1):997–1007, January 2024.
- [CCWB21] Roberto Confalonieri, Ludovik Coba, Benedikt Wagner, and Tarek R. Besold. A historical perspective of explainable artificial intelligence. *WIREs Data Mining and Knowledge Discovery*, 11(1):e1391, Jan.–Feb. 2021.
- [CD19] Marco Cavallo and Çağatay Demiralp. Clustrophile 2: Guided visual clustering analysis. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):267–276, January 2019.
- [CG19] Jesus J Caban and David Gotz. Visual analytics in healthcare – opportunities and research challenges. *Journal of the American Medical Informatics Association*, 22(2):260–262, March 2019.
- [CGL20] Zach Cutler, Kiran Gadhave, and Alexander Lex. Ttrack: A library for provenance-tracking in web-based visualizations. In *IEEE Visualization (VIS)*, Vis '20, pages 116–120, Salt Lake City, UT, October 2020.
- [CGM⁺17a] Davide Ceneda, Theresia Gschwandtner, Thorsten May, Silvia Miksch, Hans-Jörg Schulz, Marc Streit, and Christian Tominski. Amending the characterization of guidance in visual analytics. arXiv, October 2017.
- [CGM⁺17b] Davide Ceneda, Theresia Gschwandtner, Thorsten May, Silvia Miksch, Hans-Jörg Schulz, Marc Streit, and Christian Tominski. Characterizing guidance in visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):111–120, January 2017.

- [CGM⁺18] Davide Ceneda, Theresia Gschwandtner, Thorsten May, Silvia Miksch, Marc Streit, and Christian Tominski. Guidance or no guidance?: A decision tree can help. In *EuroVis Workshop on Visual Analytics*, EuroVA '18, pages 19–23, Brno, Czech Republic, 2018.
- [CGM19a] Davide Ceneda, Theresia Gschwandtner, and Silvia Miksch. A review of guidance approaches in visual data analysis: A multifocal perspective. *Computer Graphics Forum*, 38(3):861–879, June 2019.
- [CGM19b] Davide Ceneda, Theresia Gschwandtner, and Silvia Miksch. You get by with a little help: The effects of variable guidance degrees on performance and mental state. *Visual Informatics*, 3(4):177–191, December 2019.
- [CH97] Urszula Chajewska and Joseph Y. Halpern. Defining explanation in probabilistic systems. In *Proceedings of the 13th Conference on Uncertainty in Artificial Intelligence*, UAI'97, pages 62–71, Providence, RI, August 1997.
- [Chi20] Giacomo Chiesa. *Technological Paradigms and Digital Eras: Data-driven Visions for Building Design*. Springer, New York, NY, 2020.
- [CL18] Jaegul Choo and Shixia Liu. Visual analytics for explainable deep learning. *IEEE Computer Graphics and Applications*, 38(4):84–92, Jul.–Aug. 2018.
- [CLSdV22] Niltemberg de Oliveira Carvalho, Andréia Libório Sampaio, and Davi Romero de Vasconcelos. MoReXAI - a model to reason about the eXplanation design in AI systems. In *Artificial Intelligence in HCI (HCII)*, volume 13336 of *Lecture Notes in Computer Science (LNCS)*, pages 130–148, July 2022.
- [CN17] Shan Carter and Michael Nielsen. Using artificial intelligence to augment human intelligence. *Distill*, December 2017.
- [CNdS07] Fabio Campos, Andre Neves, and Fernando M. Campello de Souza. Decision making under subjective uncertainty. In *IEEE Symposium on Computational Intelligence in Multi-Criteria Decision-Making*, pages 85–90, Honolulu, HI, April 2007.
- [Cof07] Piotr Cofta. *Trust, Complexity and Control: Confidence in a Convergent World*. John Wiley & Sons, Hoboken, NJ, 2007.
- [Cor23] Visual Crossing Corporation. Weather data services | visual crossing, 2023. <https://www.visualcrossing.com/weather/weather-data-services>.
- [CPC19] Diogo V. Carvalho, Eduardo M. Pereira, and Jaime S. Cardoso. Machine learning interpretability: A survey on methods and metrics. *Electronics*, 8(8):832, July 2019.

- [CRMH12] Jason Chuang, Daniel Ramage, Christopher Manning, and Jeffrey Heer. Interpretation and trust: Designing model-driven visualizations for text analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 443–452, Austin, TX, May 2012.
- [Cui19] Wenqiang Cui. Visual analytics: A comprehensive overview. *IEEE Access*, 7:81555–81573, June 2019.
- [Cup80] Brian Cupples. Four types of explanation. *Philosophy of Science*, 47(4):626–629, December 1980.
- [CWK⁺17] Isaac Cho, Ryan Wesslen, Alireza Karduni, Sashank Santhanam, Samira Shaikh, and Wenwen Dou. The anchoring effect in decision-making with visual analytics. In *IEEE Symposium on Visual Analytics Science and Technology*, VAST '17, pages 116–126, Phoenix, AZ, October 2017.
- [CY13] Yang Chen and Jing Yang. ManyInsights: A visual analytics approach to supporting effective insight management. *Tsinghua Science and Technology*, 18(2):171–181, April 2013.
- [DAB⁺23] Alexis Delaforge, Jérôme Azé, Sandra Bringay, Caroline Mollevi, Arnaud Sallaberry, and Maximilien Servajean. EBBE-text: Explaining neural networks by exploring text classification decision boundaries. *IEEE Transactions on Visualization and Computer Graphics*, 29(10):4154–4171, October 2023.
- [DBD18] Evanthia Dimara, Anastasia Bezerianos, and Pierre Dragicevic. Conceptual and methodological issues in evaluating multidimensional visualizations for decision support. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):749–759, January 2018.
- [DFP⁺20] Evanthia Dimara, Steven Franconeri, Catherine Plaisant, Anastasia Bezerianos, and Pierre Dragicevic. A task-based taxonomy of cognitive biases for information visualization. *IEEE Transactions on Visualization and Computer Graphics*, 26(2):1413–1432, February 2020.
- [DIPJ22] Bianchi Dy, Nazim Ibrahim, Ate Poorthuis, and Sam Joyce. Improving visualization design for effective multi-objective decision making. *IEEE Transactions on Visualization and Computer Graphics*, 28(10):3405–3416, October 2022.
- [Div24] Vaghani Divyeshkumar. Balancing safety and decision-making algorithms in ai-driven autonomous vehicles. *NeuroQuantology*, 22(3):112–129, May 2024.
- [dOBD⁺22] Weiner de Oliveira, Regina Braga, José Maria N. David, Victor Ströele, Fernanda Campos, and Gabriellla Castro. Visionary: A framework for

analysis and visualization of provenance data. *Knowledge and Information Systems*, 64(2):381–413, February 2022.

- [Dod08] Yadolah Dodge. *The Concise Encyclopedia of Statistics*, chapter Exploratory Data Analysis, pages 192–194. Springer, New York, NY, 2008.
- [DPBn07] Kenneth G. DeMarree, Richard E. Petty, and Pablo Briñol. Self-certainty: Parallels to attitude certainty. *International Journal of Psychology & Psychological Therapy*, 7(2):159–188, 2007.
- [DRM19] Kimberly Dextras-Romagnino and Tamara Munzner. Segmentifier: Interactive refinement of clickstream data. *Computer Graphics Forum*, 38(3):623–634, June 2019.
- [DS22] Evanthia Dimara and John Stasko. A critical reflection on visualization research: Where do decision making tasks hide? *IEEE Transactions on Visualization and Computer Graphics*, 28(1):1128–1138, January 2022.
- [DT12] Mandeep K. Dhimi and Mary E. Thomson. On the relevance of cognitive continuum theory and quasirationality for understanding management judgment and decision making. *European Management Journal*, 30(4):316–326, August 2012.
- [dVCFP14] Ewart J. de Visser, Marvin Cohen, Amos Freedy, and Raja Parasuraman. A design methodology for trust cue calibration in cognitive agents. In *Virtual, Augmented and Mixed Reality. Designing and Developing Virtual and Augmented Environments (VAMR '14)*, volume 8525 of *Lecture Notes in Computer Science (LNCS)*, pages 251–262, Heraklion, Crete, Greece, June 2014.
- [DWH⁺22] Vaishali Dhanoa, Conny Walchshofer, Andreas Hinterreiter, Holger Stitz, Eduard Gröller, and Marc Streit. A process model for dashboard onboarding. *Computer Graphics Forum*, 41(3):501–513, June 2022.
- [EAA⁺23] Stef van den Elzen, Gennady Andrienko, Natalia Andrienko, Brian D. Fisher, Rafael M. Martins, Jaakko Peltonen, Alexandru C. Telea, and Michel Verleysen. The flow of trust: A visualization framework to externalize, explore, and explain trust in ML applications. *IEEE Computer Graphics and Applications*, 43(2):78–88, Mar.–Apr. 2023.
- [EaFB⁺94] José Encarnaç o, Jim Foley, Steve Bryson, Steven K. Feiner, and Nahum Gershon. Research issues in perception and user interfaces. *IEEE Computer Graphics and Applications*, 14(2):67–69, March 1994.
- [EAJK⁺19] Mennatallah El-Assady, Wolfgang Jentner, Rebecca Kehlbeck, Udo Schlegel, Rita Sevastjanova, Fabian Sperrle, Thilo Spinner, and Daniel A.

Keim. Towards XAI: Structuring the processes of explanations. In *Proceedings of the ACM Workshop on Human-Centered Machine Learning (HCML)*, volume 4, 2019.

- [EAM22] Mennatallah El-Assady and Caterina Moruzzi. Which biases and reasoning pitfalls do explanations trigger? decomposing communication processes in human–AI interaction. *IEEE Computer Graphics and Applications*, 42(6):11–23, Nov.–Dec. 2022.
- [Eic00] Stephen G. Eick. Visual discovery and analysis. *IEEE Transactions on Visualization and Computer Graphics*, 6(1):44–58, Jan.–Mar. 2000.
- [EK15] Ernesto Estrada and Philip A. Knight. *A First Course in Network Theory*. Oxford University Press, Oxford, England, 2015.
- [EK25] Niklas Elmqvist and Clemens Nylandsted Klokmose. Automating the path: An R&D agenda for human-centered AI and visualization. *IEEE Computer Graphics and Applications*, 45(3):73–81, May–June 2025.
- [FFT74] Mary Anne Fisherkeller, Jerome H. Friedman, and John W. Tukey. PRIM-9: An interactive multidimensional data display and analysis system. In *A.E.C. Scientific Computer Information Exchange Meeting*, May 1974.
- [FGD⁺16] Todd Fernandez, Allison Godwin, Jacqueline Doyle, Dina Verdin, Hank Boone, Adam Kirn, Lisa Benson, and Geoff Potvin. More comprehensive and inclusive approaches to demographic data collection. In *ASEE Annual Conference & Exposition*, New Orleans, LA, June 2016.
- [FGM⁺20] Katarína Furmanová, Nicolas Grossmann, Ludvig P. Muren, Oscar Casares-Magaz, Vitali Moiseenko, John P. Einck, Eduard Gröller, and Renata G. Raidou. Vapor: Visual analytics for the exploration of pelvic organ variability in radiotherapy. *Computer & Graphics*, 91:25–38, October 2020.
- [FM00] Giorgio P. Faconti and Mieke Massink. Continuity in human computer interaction. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, CHI EA '00, page 364, The Hague, The Netherlands, April 2000.
- [FMCM⁺21] Katarína Furmanová, Ludvig P. Muren, Oscar Casares-Magaz, Vitali Moiseenko, John P. Einck, Sara Pilskog, and Renata G. Raidou. PRE-VIS: Predictive visual analytics of anatomical variability for radiotherapy decision support. *Computer & Graphics*, 97:126–138, June 2021.
- [FNB⁺22] Carla Floricel, Nafiul Nipu, Mikayla Biggs, Andrew Wentzel, Guadalupe Canahuate, Lianne Van Dijk, Abdallah Mohamed, C. David Fuller, and G. Elisabeta Marai. THALIS: Human-machine analysis of longitudinal

symptoms in cancer therapy. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):151–161, January 2022.

- [FPH19] Mi Feng, Evan Peck, and Lane Harrison. Patterns and pace: Quantifying diverse exploration behavior with visualizations on the web. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):501–511, January 2019.
- [Fre79] Linton C. Freeman. Centrality in social networks conceptual clarification. *Social Networks*, 1(3):215–239, 1978-1979.
- [Fri08] Michael Friendly. *Handbook of Data Visualization*, chapter A Brief History of Data Visualization, pages 15–56. Springer, Heidelberg, Germany, 2008.
- [Gam88] Diego Gambetta. *Trust: Making and Breaking Cooperative Relations*. Basil Blackwell Ltd, Oxford, England, 1988.
- [GB09] Gerd Gigerenzer and Henry Brighton. Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science*, 1(1):107–143, January 2009.
- [GC02] Karen Gasper and Gerald L. Clore. Attending to the big picture: Mood and global versus local processing of visual information. *Psychological Science*, 13(1):34–40, January 2002.
- [GCV⁺21] Mouadh Guesmi, Mohamed Amine Chatti, Laura Vorgerd, Shoeb Joarder, Shadi Zumor, Yiqi Sun, Fangzheng Ji, and Arham Muslim. On-demand personalized explanation for transparent recommendation. In *Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*, UMAP '21, pages 246–252, Utrecht, Netherlands, June 2021.
- [GDM⁺19] Shunan Guo, Fan Du, Sana Malik, Eunyee Koh, Sungchul Kim, Zhicheng Liu, Donghyun Kim, Hongyuan Zha, and Nan Cao. Visualizing uncertainty and alternatives in event sequence predictions. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 1–12, Glasgow, Scotland, May 2019. Paper No.: 573.
- [GHG⁺19] Yolanda Gil, James Honaker, Shikhar Gupta, Yibo Ma, Vito D’Orazio, Daniel Garijo, Shruti Gadewar, Qifan Yang, and Neda Jahanshad. Towards human-guided machine learning. In *Proceedings of the International Conference on Intelligent User Interfaces*, IUI '19, pages 614–624, Marina del Ray, CA, March 2019.
- [GL12] Steven Gomez and David Laidlaw. Modeling task performance for a crowd of users from interaction histories. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 2465–2468, Austin, TX, May 2012.

- [GM21] Francesca Gauci and Maath Musleh. Agritology: A decision support system for local farmers in malta and palestine. In *CEUR Workshop Proceedings, S4BioDiv '21*, pages 1–5, Bolzano, Italy, 2021. <http://ceur-ws.org/Vol-2969/paper3-s4biodiv.pdf>.
- [GMOB22] Sneha Gathani, Shayan Monadjemi, Alvitta Ottley, and Leilani Battle. A grammar-based approach for applying visualization taxonomies to interaction logs. *Computer Graphics Forum*, 41(3):489–500, June 2022.
- [GMR⁺23] Christina Gillmann, Robin Georg Claus Maack, Felix Raith, Juan F. Pérez, and Gerik Scheuermann. A taxonomy of uncertainty events in visual analytics. *IEEE Computer Graphics and Applications*, 43(5):62–71, Sep.-Oct. 2023.
- [GMYA18] Melinda Gervasio, Karen Myers, Eric Yeh, and Boone Adkins. Explanation to avert surprise. In *Joint Proceedings of the ACM IUI Workshops*, volume 2068 of *ExSS '18*, Tokyo, Japan, March 2018. <https://ceur-ws.org/Vol-2068/exss11.pdf>.
- [Gol15] Jennifer Golbeck. *Introduction to Social Media Investigation: A Hands-on Approach*, chapter Analyzing networks, pages 221–235. Syngress, MA, 2015.
- [Gon19] Amr Mohamed Goneid. An algorithm for computing average mutual information using probability distribution smoothing. *International Journal on Recent and Innovation Trends in Computing and Communication*, 7(9):16–23, January 2019.
- [Got67] Edward E Gotts. A perceptual component of visual-analytic skills. *Head Start Evaluation and Research*, pages 31–51, 1967.
- [Gov23] Kerushan Govender. *Age of Agency*, chapter A Bit of History, pages 65–73. Productivity Press, 2023.
- [Gru24] Sebastian L. Gruener. Uncertainty decomposed: Understanding levels of contingency to enable effective decision-making. *Strategic Management*, 29(3):5–27, September 2024.
- [GSC⁺19] David Gunning, Mark Stefik, Jaesik Choi, Timothy Miller, Simone Stumpf, and Guang-Zhong Yang. XAI–explainable artificial intelligence. *Science Robotics*, 4(37), December 2019.
- [GSG⁺21] Christina Gillmann, Noeska N. Smit, Eduard Gröller, Bernhard Preim, Anna Vilanova, and Thomas Wischgoll. Ten open challenges in medical visualization. *IEEE Computer Graphics and Applications*, 41(5):7–15, Sep.-Oct. 2021.

- [GSWS21] Christina Gillmann, Dorothee Saur, Thomas Wischgoll, and Gerik Scheuermann. Uncertainty-aware visualization in medical imaging — a survey. *Computer Graphics Forum*, 40(3):665–689, June 2021.
- [GTKI11] A. Ghasemi, Mohaddeseh Taghinejad, A. Kabiri, and M. Imani. Ricoeur’s theory of interpretation: A method for understanding text (course text). *World Applied Sciences Journal*, 15(11):1623–1629, January 2011.
- [GW09] David Gotz and Zhen Wen. Behavior-driven visualization recommendation. In *Proceedings of the International Conference on Intelligent User Interfaces, IUI ’09*, pages 315–324, Sanibel Island, FL, February 2009.
- [GZ09] David Gotz and Michelle X. Zhou. Characterizing users’ visual analytic activity for insight provenance. *Information Visualization*, 8(1):42–55, January 2009.
- [Hal80] Stuart Hall. *Culture, Media, Language*, chapter Encoding/decoding, pages 163–173. Routledge, Longon, England, 1980.
- [Hal97] Stuart Hall. *Representation: Cultural Representations and Signifying Practices*, chapter The Work of Representation, pages 13–74. Sage Publications, Inc; Open University Press, Thousand Oaks, CA, 1997.
- [Han21] Wenkai Han. Making and trusting decisions in visual analytics. In *IEEE Workshop on TRust and EXPertise in Visual Analytics, TREX ’21*, pages 14–19, New Orleans, LA, October 2021.
- [Har74] Joann Harris. The computer: Guidance tool of the future. *Journal of Counseling Psychology*, 21(4):331–339, 1974.
- [HDC⁺21] Jan Hofmaier, George Dedes, David Carlson, Katia Parodi, Claus Belka, and Florian Kamp. Variance-based sensitivity analysis for uncertainties in proton therapy: A framework to assess the effect of simultaneous uncertainties in range, positioning, and RBE model predictions on RBE-weighted dose distributions. *Medical Physics*, 48(2):805–818, February 2021.
- [HEA24] Steffen Holter and Mennatallah El-Assady. Deconstructing human-AI collaboration: Agency, interaction, and adaptation. *Computer Graphics Forum*, 43(3):e15107, June 2024.
- [HEAF⁺22] Pantea Haghghatkah, Mennatallah El-Assady, Jean-Daniel Fekete, Narges Mahyar, Carita Paradis, Vasiliki Simaki, and Bettina Speckmann. Characterizing uncertainty in the visual text analysis pipeline. In *IEEE 7th Workshop on Visualization for the Digital Humanities, VIS4DH ’22*, pages 25–30, Oklahoma City, OK, October 2022.

- [HEF⁺18] Trung Dong Huynh, Mark Ebden, Joel Fischer, Stephen Roberts, and Luc Moreau. Provenance network analytics: An approach to data analytics using data provenance. *Data Mining and Knowledge Discovery*, 23:708–735, February 2018.
- [Hev18] David Hevey. Network analysis: a brief overview and tutorial. *Health Psychology and Behavioral Medicine*, 6(1):301–328, September 2018.
- [HHH⁺11] Leanne M. Hirshfield, Stuart H. Hirshfield, Samuel Hincks, Matthew Russell, Rachel Ward, and Tom Williams. Trust in human-computer interactions as measured by frustration, surprise, and workload. In *Foundations of Augmented Cognition. Directing the Future of Adaptive Systems (FAC)*, volume 6780 of *Lecture Notes in Computer Science (LNCS)*, pages 507–516, Orlando, FL, July 2011.
- [HJCM22] Emily Hindalong, Jordon Johnson, Giuseppe Carenini, and Tamara Munzner. Abstractions for visualizing preferences in group decisions. In *Proceedings of the ACM on Human-Computer Interaction*, volume 6(CSCW1), New Orleans, LA, April 2022. Article no. 49.
- [HKPC19] Fred Hohman, Minsuk Kahng, Robert Pienta, and Duen Horng Chau. Visual analytics in deep learning: An interrogative survey for the next frontiers. *IEEE Transactions on Visualization and Computer Graphics*, 25(8):2674–2693, August 2019.
- [HKW23] David Hägele, Tim Krake, and Daniel Weiskopf. Uncertainty-aware multi-dimensional scaling. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):23–32, January 2023.
- [HLS⁺12] Clemens Holzhüter, Alexander Lex, Dieter Schmalstieg, Hans-Jörg Schulz, Heidrun Schumann, and Marc Streit. Visualizing uncertainty in biological expression data. In *Proceedings of the SPIE Conference on Visualization and Data Analysis*, volume 8294, pages 1–11, January 2012.
- [HMH⁺21] Chen He, Luana Micallef, Liye He, Gopal Peddinti, Tero Aittokallio, and Giulio Jacucci. Characterizing the quality of insight by interactions: A case study. *IEEE Transactions on Visualization and Computer Graphics*, 27(8):3410–3424, August 2021.
- [HMKB23] Jinbin Huang, Aditi Mishra, Bum Chul Kwon, and Chris Bryan. Conceptexplainer: Interactive explanation for deep neural networks from a concept perspective. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):831–841, January 2023.
- [HMO24] Sunwoo Ha, Shayan Monadjemi, and Alvitta Ottley. Guided by AI: Navigating trust, bias, and data exploration in AI-guided visual analytics. *Computer Graphics Forum*, 43(3):e15108, June 2024.

- [HN98] Jerry L. Hintze and Ray D. Nelson. Violin plots: A box plot-density trace synergism. *The American Statistician*, 52(2):181–184, March 1998.
- [HQC⁺19] Jessica Hullman, Xiaoli Qiao, Michael Correll, Alex Kale, and Matthew Kay. In pursuit of error: A survey of uncertainty visualization evaluation. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):903–913, January 2019.
- [HR19] Lars Kai Hansen and Laura Rieger. *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, chapter Interpretability in Intelligent Systems—A New Concept?, pages 41–49. Springer, 2019.
- [HS23] Wenkai Han and Hans-Jörg Schulz. Providing visual analytics guidance through decision support. *Information Visualization*, 22(2):140–165, January 2023.
- [HSB⁺22] David Hägele, Christoph Schulz, Cedric Beschle, Hannah Booth, Miriam Butt, Andrea Barth, Oliver Deussen, and Daniel Weiskopf. Uncertainty visualization: Fundamentals and recent developments. *it - Information Technology*, 64(4-5):121–132, August 2022.
- [HSE24] Md Naimul Hoque, Sungbok Shin, and Niklas Elmqvist. Harder, better, faster, stronger: Interactive visualization for human-centered AI tools. arXiv, November 2024.
- [HV08] James S. Hernandez and Prathibha Varkey. Vertical versus lateral thinking. *Physician Executive*, 34(3):26–28, May-Jun. 2008.
- [HZW21] Ambreen Hanif, Xuyun Zhang, and Steven Wood. A survey on explainable artificial intelligence techniques and challenges. In *IEEE 25th International Enterprise Distributed Object Computing Workshop (EDOCW)*, pages 81–89, Gold Coast, Australia, October 2021.
- [Ibn09] Ibn Tufayl. *Hay Ibn Yaqzān*. Dar al-Afaq al-Jadida, Beirut, Lebanon, 2009.
- [IIC⁺13] Tobias Isenberg, Petra Isenberg, Jian Chen, Michael Sedlmair, and Torsten Möller. A systematic review on the practice of evaluating visualization. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2818–2827, December 2013.
- [Jau22] Théo Jaunet. *Deep Learning Interpretability with Visual Analytics: Exploring Reasoning and Bias Exploitation*. Thesis, Université de Lyon, May 2022.
- [JED⁺20] Amit Jena, Ulrich Engelke, Tim Dwyer, Venkatesh Raiamanickam, and Cecile Paris. Uncertainty visualisation: An interactive visual survey.

In *IEEE Pacific Visualization Symposium (Pacific Vis)*, pages 201–205, Tianjin, China, June 2020.

- [Jon12] Karen Jones. Trustworthiness. *Ethics*, 123(1):61–85, October 2012.
- [JS22] Mladen Jovanović and Mia Schmitz. Explainability as a user requirement for artificial intelligence systems. *Computer*, 55(2):90–94, February 2022.
- [KAF⁺08] Daniel Keim, Gennady Andrienko, Jean-Daniel Fekete, Carsten Görg, Jörn Kohlhammer, and Guy Melançon. *Information Visualization: Human-Centered Issues and Perspectives*, volume 4950 of *Lecture Notes in Computer Science (LNCS)*, chapter Visual Analytics: Definition, Process, and Challenges, pages 154–175. Springer, Heidelberg, 2008.
- [KDJ⁺21] Aasim Kamal, Parashar Dhakal, Ahmad Y. Javaid, Vijay K. Devabhaktuni, Devinder Kaur, Jack Zaiantz, and Robert Marinier. Recent advances and challenges in uncertainty visualization: A survey. *Journal of Visualization*, 24(5):861–890, May 2021.
- [Ker22] Trish Kerin. Which eye are you looking through? the impact of cognitive bias on process safety. *Process Safety Progress*, 41(1):101–105, March 2022.
- [KH79] Peter G. W. Keen and Richard D. Hackathorn. *Decision support systems and personal computing*. MIT: Center for Information Systems Research, Cambridge, MA, October 1979.
- [Kha13] Kareem Khalifa. The role of explanation in understanding. *The British Journal for the Philosophy of Science*, 64(1):161–187, March 2013.
- [KJV⁺79] Hans J. Keim, Philip M. Johnson, E. Darracott Vaughan, Jr., Khalid Beg, David A. Follett, Leonard M. Freeman, and John H. Laragh. Computer-assisted static/dynamic renal imaging: A screening test for renovascular hypertension? *Journal of Nuclear Medicine*, 20(1):11–17, 1979.
- [KK17] Natalie Kerracher and Jessie Kennedy. Constructing and evaluating visualisation task classifications: Process and considerations. *Computer Graphics Forum*, 36(3):47–59, June 2017.
- [KKP⁺11] Jörn Kohlhammer, Daniel Keim, Margit Pohl, Giuseppe Santucci, and Gennady Andrienko. Solving problems with visual analytics. *Procedia Computer Science*, 7:117–120, 2011. Proceedings of the 2nd European Future Technologies Conference and Exhibition 2011 (FET 11).
- [KLL⁺22] Xiaoyan Kui, Naiming Liu, Qiang Liu, Jingwei Liu, Xiaoqian Zeng, and Chao Zhang. A survey of visual analytics techniques for online education. *Visual Informatics*, 6(4):67–77, December 2022.

- [KMT10] Daniel A. Keim, Florian Mansmann, and Jim Thomas. Visual analytics: how much visualization and how much analytics? *ACM SIGKDD Explorations Newsletter*, 11(2):5–8, May 2010.
- [Kor12] George Kormoz. *The Guide to Organic Gardening*. Ma'an Development Center, Gaza, Palestine, 2012.
- [KPBG13] Johannes Kehrer, Harald Piringer, Wolfgang Berger, and M. Eduard Gröller. A model for structure-based comparison of many categories in small-multiple displays. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2287–2296, December 2013.
- [KSD⁺22] Kostiantyn Kucher, Nicole Sultanum, Angel Daza, Vasiliki Simaki, Maria Skeppstedt, Barbara Plank, Jean-Daniel Fekete, and Narges Mahyar. An interdisciplinary perspective on evaluation and experimental design for visual text analytics: Position paper. In *IEEE Evaluation and Beyond - Methodological Approaches for Visualization*, BELIV '22, pages 28–37, Oklahoma City, OK, October 2022.
- [KW19] Florian Kamp and Jan J. Wilkens. Application of variance-based uncertainty and sensitivity analysis to biological modeling in carbon ion treatment plans. *Medical Physics*, 46(2):437–447, February 2019.
- [LBI⁺12] Heidi Lam, Enrico Bertini, Petra Isenberg, Catherine Plaisant, and Sheelagh Carpendale. Empirical studies in information visualization: Seven scenarios. *IEEE Transactions on Visualization and Computer Graphics*, 18(9):1520–1536, September 2012.
- [LBS⁺19] Timothy Luciani, Andrew Burks, Cassiano Sugiyama, Jonathan Komperda, and G. Elisabeta Marai. Details-first, show context, overview last: Supporting exploration of viscous fingers in large-scale ensemble simulations. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):1225–1235, January 2019.
- [LF13] Q. Vera Liao and Wai-Tat Fu. Beyond the filter bubble: Interactive effects of perceived threat and topic involvement on selective exposure to information. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 2359–2368, Paris, France, April 2013.
- [LRBB⁺23] B. La Rosa, G. Blasilli, R. Bourqui, D. Auber, G. Santucci, R. Capobianco, E. Bertini, R. Giot, and M. Angelini. State of the art of visual analytics for eXplainable deep learning. *Computer Graphics Forum*, 42(1):319–355, February 2023.

- [LS15] Hyunjoung Lee and Il Sohn. *Fundamentals of Big Data Network Analysis for Research and Industry*. John Wiley & Sons, West Sussex, England, 2015.
- [Luh05] Niklas Luhmann. *Risk: A Sociological Risk*. Routledge, New York, NY, 2005.
- [Lux25] Luxembourg Institute of Science and Technology (LIST). PROGRAM EuroVis2025, 2025. <https://www.eurovis2025.lu/for-attendees/program>.
- [LW90] Donna M. Lamberti and William A. Wallace. Intelligent interface design: An empirical assessment of knowledge presentation in expert systems. *MIS Quarterly*, 14(3):279–311, September 1990.
- [MA14] Silvia Miksch and Wolfgang Aigner. A matter of time: Applying a data–users–tasks design triangle to visual analytics of time-oriented data. *Computer & Graphics*, 38:286–290, February 2014.
- [Mac15] Alan M. MacEachren. Visual analytics and uncertainty: Its not about the data. In *EuroVis Workshop on Visual Analytics*, EuroVA '15, Cagliari, Sardinia, May 2015.
- [MAK⁺23] J. Méndez, C. Alrabbaa, P. Koopmann, R. Langner, F. Baader, and R. Dachzelt. Evonne: A visual tool for explaining reasoning with OWL ontologies and supporting interactive debugging. *Computer Graphics Forum*, 42(6):e14730, September 2023.
- [MBBM22] D. Morrison, M. Bedinger, L. Beevers, and K. McClymont. Exploring the raison d’être behind metric selection in network analysis: a systematic review. *Applied Network Science*, 7(1):50, July 2022.
- [MCER25] Maath Musleh, Davide Ceneda, Henry Ehlers, and Renata G. Raidou. ConAn: Measuring and evaluating user confidence in visual data analysis under uncertainty. *Computer Graphics Forum*, 44(1):e15272, February 2025.
- [MCJ22] Maath Musleh, Angelos Chatzimparmpas, and Ilir Jusufi. Visual analysis of blow molding machine multivariate time series data. *Journal of Visualization*, 25:1329–1342, July 2022.
- [McK80] Robert H. McKim. *Experiences in Visual Thinking*. PWS Publishing, Boston, MA, second edition, 1980.
- [MD19] Miriah Meyer and Jason Dykes. Criteria for rigor in visualization design study. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):87–97, January 2019.

- [MGG⁺23] Shayan Monadjemi, Mengtian Guo, David Gotz, Roman Garnett, and Alvitta Ottley. Human–computer collaboration for visual analytics: an agent-based framework. *Computer Graphics Forum*, 42(3):199–210, June 2023.
- [MHTD24] L. E. Matzen, B. C. Howell, M. Tuft, and K. M. Divis. Transparent risks: The impact of the specificity and visual encoding of uncertainty on decision making. *Computer Graphics Forum*, 43(3), June 2024.
- [Mil19] Tim Miller. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267:1–38, February 2019.
- [MK83] Masahiro Mizuta and Michiaki Kawaguchi. A graphical dynamic representation of multi-dimensional data. *Journal of the Japan Statistical Society*, 13(1):1–10, 1983.
- [MKLS13] Suzanne Morony, Sabina Kleitman, Yim Ping Lee, and Lazar Stankov. Predicting achievement: Confidence vs self-efficacy, anxiety, and self-concept in confucian and european countries. *International Journal of Educational Research*, 58:79–96, 2013.
- [MLC20] Silvia Miksch, Heike Leitte, and Min Chen. *Foundations of Data Visualization*, chapter Knowledge-Assisted Visualization and Guidance, pages 61–85. Springer, New York, NY, August 2020.
- [MMT⁺23] Maath Musleh, Ludvig Paul Muren, Laura Toussaint, Anne Vestergaard, Eduard Gröller, and Renata G. Raidou. Uncertainty guidance in proton therapy planning visualization. *Computer & Graphics*, 111:166–179, April 2023.
- [Moo16] James W. Moore. What is the sense of agency and why does it matter? *Frontiers in Psychology*, 7:1272, August 2016.
- [MPG⁺17] Radhe Mohan, Christopher R. Peeler, Fada Guan, Lawrence Bronk, Wenhua Cao, and David R. Grosshans. Radiobiological issues in proton therapy. *Acta Oncologica*, 56(11):1367–1373, November 2017.
- [MRC25] Maath Musleh, Renata G. Raidou, and Davide Ceneda. Trustme: A context-aware explainability model to promote user trust in guidance. *IEEE Transactions on Visualization and Computer Graphics*, 31(10):8040–8056, October 2025.
- [MRP⁺25] Robin G. C. Maack, Felix Raith, Juan F. Pérez, Gerik Scheuermann, and Christina Gillmann. A workflow to systematically design uncertainty-aware visual analytics applications. *The Visual Computer*, 41(3):1485–1498, June 2025.

- [MSH⁺23] Robin G. C. Maack, Gerik Scheuermann, Hans Hagen, Jose Tiberio Hernández Peñaloza, and Christina Gillmann. Uncertainty-aware visual analytics: scope, opportunities, and challenges. *The Visual Computer*, 39:6345–6366, December 2023.
- [Mui87] Bonnie M. Muir. Trust between humans and machines, and the design of decision aids. *International Journal of Man-Machine Studies*, 27(5-6):527–539, November 1987.
- [Mul70] Mervin E. Muller. Computers as an instrument for data analysis. *Technometrics*, 12(2):259–293, May 1970.
- [Mus24a] Maath Musleh. ConAn. OSF, December 2024. <http://doi.org/10.17605/OSF.IO/24HFN>.
- [Mus24b] Maath Musleh. TrustMe. OSF, December 2024. <http://doi.org/10.17605/OSF.IO/YT2F7>.
- [Mus25] Maath Musleh. Dissertation. OSF, September 2025. <http://doi.org/10.17605/OSF.IO/2SHXC>.
- [MVH⁺21] Shane T. Mueller, Elizabeth S. Veinott, Robert R. Hoffman, Gary Klein, Lamia Alam, Tauseef Mamun, and William J. Clancey. Principles of explanation in human-AI systems. arXiv, February 2021.
- [MVM23] Aisha Muthaffar and Sonia Vilches-Montero. Empowering retailers: A bounded rationality perspective to enhancing omnichannel journey satisfaction. *Journal of Retailing and Consumer Services*, 75:103536, November 2023.
- [MW09] Samantha Meyer and Paul Ward. Reworking the sociology of trust: Making a semantic distinction between trust and dependence. In *The Future of Sociology. Proceedings of the Australian Sociological Association Conference*, pages 1–16, 2009.
- [MZR21] Sina Mohseni, Niloofar Zarei, and Eric D. Ragan. A multidisciplinary survey and framework for design and evaluation of explainable AI systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 11(3–4):1–45, 2021. Article no. 24.
- [NCWE22] Arpit Narechania, Adam Coscia, Emily Wall, and Alex Endert. Lumos: Increasing awareness of analytic behavior during visual data analysis. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):1009–1018, January 2022.
- [NSS⁺24] Kawa Nazemi, Cristian A. Secco, Lennart B. Sina, Uliana Eliseeva, Elena Correll, and Midhad Blazevic. Visual analytics for decision-making. In *28th*

International Conference Information Visualisation (IV), pages 150–159, Coimbra, Portugal, July 2024.

- [NTA⁺19] Phong H. Nguyen, Cagatay Turkay, Gennady Andrienko, Natalia Andrienko, Olivier Thonnard, and Jihane Zouaoui. Understanding user behaviour through action sequences: From the usual to the unusual. *IEEE Transactions on Visualization and Computer Graphics*, 25(9):2838–2852, September 2019.
- [NWC⁺21] Carolina Nobre, Dylan Wootton, Zach Cutler, Lane Harrison, Hanspeter Pfister, and Alexander Lex. reVISit: Looking under the hood of interactive visualization studies. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI '21, pages 1–13, Yokohama, Japan, May 2021. Article no. 25.
- [NWHL20] Carolina Nobre, Dylan Wootton, Lane Harrison, and Alexander Lex. Evaluating multivariate network visualization techniques using a validated design and crowdsourcing approach. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI '20, pages 1–12, Honolulu, HI, April 2020.
- [NXB⁺16] Phong H. Nguyen, Kai Xu, Andy Bardill, Betul Salman, Kate Herd, and B.L. William Wong. Sensemap: Supporting browser-based online sensemaking through analytic provenance. In *IEEE Symposium on Visual Analytics Science and Technology*, VAST '16, pages 91–100, Baltimore, MD, October 2016.
- [OAB⁺17] Weiner Oliveira, Lenitta M. Ambrósio, Regina Braga, Victor Ströele, José Maria David, and Fernanda Campos. A framework for provenance analysis and visualization. *Procedia Computer Science*, 108:1592–1601, 2017.
- [OCW⁺24] Emre Oral, Ria Chawla, Michel Wijkstra, Narges Mahyar, and Evanthia Dimara. From information to choice: A critical inquiry into visualization tools for decision making. *IEEE Transactions on Visualization and Computer Graphics*, 30(1):359–369, January 2024.
- [OIT⁺21] Ole Marius Otterlei, Daniel J. Indelicato, Laura Toussaint, Kristian S. Ytre-Hauge, Sara Pilskog, Lars Fredrik Fjaera, Eivind Rørvik, Helge Egil S. Pettersen, Ludvig P. Muren, Yasmin Lassen-Ramshad, Marcos Di Pinto, and Camilla H. Stokkevåg. Variation in relative biological effectiveness for cognitive structures in proton therapy of pediatric brain tumors. *Acta Oncologica*, 60(2):267–274, 2021.
- [OLA19] Richard Osuala, Jieyi Li, and Ognjen Arandjelovic. Bringing modern machine learning into clinical practice through the use of intuitive visual-

ization and human-computer interaction. *Augmented Human Research*, 4:3, 2019.

- [Om 18] Om Sleiman Farm. *Trainees Essentials: Training Resistance Farming Program*. Om Sleiman Farm, Bilin, Palestine, 2018.
- [OV21] Jeroen Ooge and Katrien Verbert. Trust in prediction models: a mixed-methods pilot study on the impact of domain expertise. In *IEEE Workshop on TRust and EXpertise in Visual Analytics*, TREX '21, pages 8–13, New Orleans, LA, October 2021.
- [OWM08] Rick O’Gorman, David Sloan Wilson, and Ralph R. Miller. An evolved cognitive bias for social norms. *Evolution and Human Behavior*, 29(2):71–78, 2008.
- [Pag18] Harald Paganetti. Proton relative biological effectiveness - uncertainties and opportunities. *International Journal of Particle Therapy*, 5(1):2–14, June 2018.
- [Pal23] Palestinian Central Bureau of Statistics. PCBS | Consumer Price Index, 2023. <https://www.pcbs.gov.ps>.
- [PCRHS18] Lace M. Padilla, Sarah H. Creem-Regehr, Mary Hegarty, and Jeanine K. Stefanucci. Decision making with visualizations: a cognitive framework across disciplines. *Cognitive Research: Principles and Implications*, 3:29, July 2018.
- [PFCB23] Lace Padilla, Racquel Fygenon, Spencer C. Castro, and Enrico Bertini. Multiple forecast visualizations (MFVs): Trade-offs in trust and performance in multiple COVID-19 forecast visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):12–22, January 2023.
- [PFW⁺18] Vanessa Peña-Araya, Consuelo Martínez Fontaine, Xiang Wei, Guillaume Delpech, and Anastasia Bezerianos. Uncertainty in science is malleable. advocating for user-agency in defining uncertainty in visualizations: a case study in geology. In *EuroVis Workshop on Reproducibility, Verification, and Validation in Visualization (EuroRV3)*, Brno, Czech Republic, June 2018.
- [PG15] Pamela Paxton and Jennifer L. Glanville. Is trust rigid or malleable? a laboratory experiment. *Social Psychology Quarterly*, 78(2):194–204, June 2015.
- [PGA13] Kristin Potter, Samuel Gerber, and Erik W. Anderson. Visualization of uncertainty without a mean. *IEEE Computer Graphics and Applications*, 33(1):75–79, Jan.-Feb. 2013.

- [PKH21] Lace Padilla, Matthew Kay, and Jessica Hullman. *Wiley StatsRef: Statistics Reference Online*, chapter Uncertainty Visualization, pages 1–18. John Wiley & Sons, February 2021.
- [PL20] Bernhard Preim and Kai Lawonn. A survey of visual analytics for public health. *Computer Graphics Forum*, 39(1):543–580, February 2020.
- [PM22] Georgia Panagiotidou and Andrew Vande Moere. Communicating qualitative uncertainty in data visualization: Two cases from within the digital humanities. *Information Design Journal*, 27(1):52–63, December 2022.
- [PMCEA⁺22] I. Pérez-Messina, D. Ceneda, M. El-Assady, S. Miksch, and F. Sperrle. A typology of guidance tasks in mixed-initiative visual analytics environments. *Computer Graphics Forum*, 41(3):465–476, June 2022.
- [PMCM24] Ignacio Pérez-Messina, Davide Ceneda, and Silvia Miksch. Guided visual analytics for image selection in time and space. *IEEE Transactions on Visualization and Computer Graphics*, 30(1):66–75, January 2024.
- [Poh17] Rüdiger F. Pohl. *Cognitive Illusions: Intriguing Phenomena in Thinking, Judgment, and Memory*. Routledge, Oxon, England, third edition, 2017.
- [Pol83] Neil W. Polhemus. Statistical graphics in APL: Structuring a graphical support system for interactive data analysis. *Computer & Graphics*, 7(1):73–80, 1983.
- [Pre76] K. Preston Jr. *Digital Processing of Biomedical Images*, chapter Digital Image Processing in the United States, pages 1–10. Springer, Boston, MA, 1976.
- [PRJ⁺23] Ameya Patil, Gaëlle Richer, Christopher Jermaine, Dominik Moritz, and Jean-Daniel Fekete. Studying early decision making with progressive bar charts. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):407–417, January 2023.
- [PRSL23] Bernhard Preim, Renata Raidou, Noeska Smit, and Kai Lawonn. *Visualization, Visual Analytics and Virtual Reality in Medicine: State-of-the-Art Techniques and Applications*. Elsevier Science, London, England, 2023.
- [PSBn23] Richard E. Petty, Joseph J. Siev, and Pablo Briñol. Attitude strength: What’s new? *The Spanish Journal of Psychology*, 26:e4, 2023.
- [PVPM22] Georgia Panagiotidou, Ralf Vandam, Jeroen Poblome, and Andrew Vande Moere. Implicit error, uncertainty and confidence in visualization: An archaeological case study. *IEEE Transactions on Visualization and Computer Graphics*, 28(12):4389–4402, December 2022.

- [PW06] Matthew D. Plumlee and Colin Ware. Zooming versus multiple window interfaces: Cognitive costs of visual comparisons. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 13(2):179–209, June 2006.
- [PWL97] Alex T. Pang, Craig M. Wittenbrink, and Suresh K. Lodha. Approaches to uncertainty visualization. *The Visual Computer*, 13:370–390, November 1997.
- [QPCRM15] P. Samuel Quinan, Lace Padilla, Sarah H. Creem-Regehr, and Miriah Meyer. Towards ecological validity in evaluating uncertainty. In *Proceedings of the IEEE VIS Workshop on Visualization for Decision Making Under Uncertainty*, pages 1–2, October 2015.
- [QZC22] Jianhong Qu, Ronggang Zhou, and Zhe Chen. The effect of personal pronouns on users and the social role of conversational agents. *Behaviour & Information Technology*, 41(16):3470–3486, November 2022.
- [Rai18] Renata G. Raidou. Uncertainty visualization: Recent developments and future challenges in prostate cancer radiotherapy planning. In *EuroVis Workshop on Reproducibility, Verification, and Validation in Visualization (EuroRV3)*, Brno, Czech Republic, June 2018.
- [RBCR17] Devarajan Ramanujan, William Z. Bernstein, Senthil K. Chandrasegaran, and Karthik Ramani. Visual analytics tools for sustainable lifecycle design: Current status, challenges, and future opportunities. *Journal of Mechanical Design*, 139(11):111415, November 2017.
- [RBMK24] Pavlo Radiuk, Olexander Barmak, Eduard Manziuk, and Iurii Krak. Explainable deep learning: A visual analytics approach with transition matrices. *Mathematics*, 12(7):1024, March 2024.
- [RBS⁺18] Marc Rautenhaus, Michael Böttinger, Stephan Siemen, Robert Hoffman, Robert M. Kirby, Mahsa Mirzargar, Niklas Röber, and Rüdiger Westermann. Visualization in meteorology—a survey of techniques and tools for data analysis tasks. *IEEE Transactions on Visualization and Computer Graphics*, 24(12):3268–3296, December 2018.
- [RESC16] Eric D. Ragan, Alex Endert, Jibonananda Sanyal, and Jian Chen. Characterizing provenance in visualization and data analysis: An organizational framework of provenance types and purposes. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):31–40, January 2016.
- [RFD⁺18] Eivind Rørvik, Lars Fredrik Fjæra, Tordis J Dahle, Jon Espen Dale, Grete May Engeseth, Camilla H Stokkevåg, Sara Thörnqvist, and Kristian S Ytre-Hauge. Exploration and application of phenomenological RBE models for proton therapy. *Physics in Medicine and Biology*, 63(18):185013, September 2018.

- [RFG⁺20] Renata Georgia Raidou, Katarína Furmanová, Nicolas Grossmann, Oscar Casares-Magaz, Vitali Moiseenko, John P. Einck, Eduard Gröller, and Ludvig P. Muren. Lessons learnt from developing visual analytics applications for adaptive prostate cancer radiotherapy. In *The Gap between Visualization Research and Visualization Software (VisGap)*, pages 51–58, Norrköping, Sweden, May 2020.
- [RLN07] Ruth Rosenholtz, Yuanzhen Li, and Lisa Nakano. Measuring visual clutter. *Journal of Vision*, 7(2):17, August 2007.
- [Rob96] Sandra L. Robinson. Trust and breach of the psychological contract. *Administrative Science Quarterly*, 41(4):574–599, December 1996.
- [RPHL14] Gordan Ristovski, Tobias Preusser, Horst K. Hahn, and Lars Linsen. Uncertainty in medical visualization: Towards a taxonomy. *Computer & Graphics*, 39:60–73, April 2014.
- [SAM⁺23] Jordan Richard Schoenherr, Roba Abbas, Katina Michael, Pablo Rivas, and Theresa Dirndorfer Anderson. Designing AI using a human-centered approach: Explainability and accuracy toward trustworthiness. *IEEE Transactions on Technology & Society*, 4(1):9–23, March 2023.
- [SC20] Ramya Srinivasan and Ajay Chander. Explanation perspectives from the cognitive sciences—a survey. In *Proceedings of the 29th International Joint Conference on Artificial Intelligence, IJCAI-20*, pages 4812–4818, 2020.
- [SCEA23] Fabian Sperrle, Davide Ceneda, and Mennatallah El-Assady. Lotse: A practical framework for guidance in visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):1124–1134, January 2023.
- [SCW⁺22] Christina Stoiber, Davide Ceneda, Markus Wagner, Victor Schetinger, Theresia Gschwandtner, Marc Streit, Silvia Miksch, and Wolfgang Aigner. Perspectives of visualization onboarding and guidance in VA. *Visual Informatics*, 6(1):68–83, March 2022.
- [SEAAC24] Fabian Sperrle, Mennatallah El-Assady, Alessio Arleo, and Davide Ceneda. A wizard of oz study of guidance strategies and dynamics. *IEEE Transactions on Visualization and Computer Graphics*, pages 1–15, July 2024. Early Access.
- [SF20] Kacper Sokol and Peter Flach. Explainability fact sheets: A framework for systematic assessment of explainable approaches. In *Proceedings of the Conference on Fairness, Accountability, and Transparency, FAT* '20*, pages 56–67, Barcelona, Spain, January 2020.
- [SGRn09] Emilio Sánchez, Héctor García-Rodicio, and Santiago R. Acuña. Are instructional explanations more effective in the context of an impasse? *Instructional Science*, 37:537–563, 2009.

- [SHK16] Joshua I. Sanders, Balázs Hangya, and Adam Kepecs. Signatures of a statistical computation in the human sense of confidence. *Neuron*, 90(3):499–506, May 2016.
- [Shn22] Ben Shneiderman. *Human-centered AI*. Oxford University Press, Oxford, England, 2022.
- [SIBB11] Michael Sedlmair, Petra Isenberg, Dominikus Baur, and Andreas Butz. Information visualization evaluation in large companies: Challenges, experiences and recommendations. *Information Visualization*, 10(3):248–266, July 2011.
- [Sil91] Mark S. Silver. Decisional guidance for computer-based decision support. *MIS Quarterly*, 15(1):105–122, March 1991.
- [Sim07] Jeffrey A. Simpson. Psychological foundations of trust. *Current Directions in Psychological Science*, 16(5):264–268, October 2007.
- [SJ24] Martin Svensson and Mattias Jacobsson. Managing inconsistencies in medical decision-making: An eight-fold typology. *European Management Journal*, 42(1):130–141, February 2024.
- [SJB⁺20] Fabian Sperrle, Astrik Jeitler, Jürgen Bernard, Daniel A. Keim, and Mennatallah El-Assady. Learning and teaching in co-adaptive guidance for mixed-initiative visual analytics. In *EuroVis Workshop on Visual Analytics*, EuroVA '20, Norrköping, Sweden, May 2020.
- [SJB⁺21] Fabian Sperrle, Astrik Jeitler, Jürgen Bernard, Daniel Keim, and Mennatallah El-Assady. Co-adaptive visual data analysis and guidance processes. *Computer & Graphics*, 100:93–105, November 2021.
- [SKJ15] Lazar Stankov, Sabina Kleitman, and Simon A. Jackson. Chapter 7 - measures of the trait of confidence. In *Measures of Personality and Social Psychological Constructs*, pages 158–189. Academic Press, San Diego, CA, 2015.
- [SLB⁺18] Cheng B. Saw, SiCong Li, Frank Battin, Janice McKeague, and Christopher A. Peters. External beam planning module of eclipse for external beam radiation therapy. *Medical Dosimetry*, 43(2):195–204, Summer 2018.
- [SLMK18] Jim Smith, Phil Legg, Milos Matovic, and Kristofer Kinsey. Predicting user confidence during visual decision making. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 8(2):10, June 2018.
- [SLSG16] H. Stitz, S. Luger, M. Streit, and N. Gehlenborg. AVOCADO: Visualization of workflow-derived data provenance for reproducible biomedical research. *Computer Graphics Forum*, 35(3):481–490, June 2016.

- [SM19] Wojciech Samek and Klaus-Robert Müller. *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, volume 11700 of *Lecture Notes in Computer Science (LNCS)*, chapter Towards Explainable Artificial Intelligence, pages 5–22. Springer, 2019.
- [SME08] Anya Savikhin, Ross Maciejewski, and David S. Ebert. Applied visual analytics for economic decision-making. In *IEEE Symposium on Visual Analytics Science and Technology, VAST '08*, pages 107–114, Columbus, OH, October 2008.
- [Smi05] Carole Smith. Understanding trust and confidence: Two paradigms and their significance for health and social care. *Journal of Applied Philosophy*, 22(3):299–316, November 2005.
- [SMM12] Michael Sedlmair, Miriah Meyer, and Tamara Munzner. Design study methodology: Reflections from the trenches and the stacks. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2431–2440, December 2012.
- [SPA23] Christina Stoiber, Margit Pohl, and Wolfgang Aigner. Design actions for the design of visualization onboarding methods. In *IEEE VIS Workshop on Visualization Education, Literacy, and Activities (EduVis)*, pages 1–10, Melbourne, Australia, October 2023.
- [SPBR20] Matthias Schlachter, Bernhard Preim, Katja Bühler, and Renata G. Raidou. Principles of visualization in radiation oncology. *Oncology*, 98(6):412–422, June 2020.
- [Spe07] Robert Spence. *Information Visualization: Design for Interaction*. Pearson Education, Essex, England, second edition, 2007.
- [SRM⁺19] M. Schlachter, R. G. Raidou, L. P. Muren, B. Preim, P.M. Putora, and K. Bühler. State-of-the-art report: Visual computing in radiation therapy planning. *Computer Graphics Forum*, 38(3):753–779, June 2019.
- [SSK⁺16] Dominik Sacha, Hansi Senaratne, Bum Chul Kwon, Geoffrey Ellis, and Daniel A. Keim. The role of uncertainty, awareness, and trust in visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):240–249, January 2016.
- [SSS⁺14] Dominik Sacha, Andreas Stoffel, Florian Stoffel, Bum Chul Kwon, Geoffrey Ellis, and Daniel A. Keim. Knowledge generation model for visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1604–1613, December 2014.
- [SSSEA20] Thilo Spinner, Udo Schlegel, Hanna Schäfer, and Mennatallah El-Assady. explAIner: A visual analytics framework for interactive and explainable

machine learning. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):1064–1074, January 2020.

- [Sta14] John Stasko. Value-driven evaluation of visualizations. In *Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*, BELIV '14, pages 46–53, Paris, France, November 2014.
- [STS⁺12] Junji Suzuki, Kunihiko Tateoka, Katsumi Shima, Yuji Yaegashi, Kazunori Fujimoto, Yuichi Saitoh, Akihiro Nakata, Tadanori Abe, Takuya Nakazawa, Kouichi Sakata, and Masato Hareyama. Uncertainty in patient set-up margin analysis in radiation therapy. *Journal of Radiation Research*, 53(4):615–619, July 2012.
- [SV23] Francesco Sovrano and Fabio Vitali. An objective metric for explainable AI: How and why to estimate the degree of explainability. *Knowledge-Based Systems*, 278:110866, October 2023.
- [SWG⁺21] Christina Stoiber, Conny Walchshofer, Florian Grassinger, Holger Stitz, Marc Streit, and Wolfgang Aigner. Design and comparative evaluation of visualization onboarding methods. In *The 14th International Symposium on Visual Information Communication and Interaction*, VINCI '21, pages 1–5, Potsdam, Germany, September 2021.
- [TCK⁺01] Jim Thomas, Paula Cowley, Olga Kuchar, Lucy Nowell, Judi Thomson, and Pak Chung Wong. Discovering knowledge through visual analysis. *Journal of Universal Computer Science*, 7(6):517–529, June 2001.
- [TF22] Fatima Tfaily and Mohamad M. Fouad. Multi-level stacking of lstm recurrent models for predicting stock-market indices. *Data Science in Finance and Economics*, 2(2):147–162, May 2022.
- [TP21] Roberto Theron and Lace M. Padilla. Editorial: Uncertainty visualization and decision making. *Frontiers in Computer Science*, 3, September 2021.
- [Tro02] J. D. Trout. Scientific explanation and the sense of understanding. *Philosophy of Science*, 69(2):212–233, 2002.
- [TTvE16] Alexander Toet, Susanne Tak, and Johannes Bernardus Fransiscus van Erp. Visualizing uncertainty. towards a better understanding of weather forecasts. *Tijdschrift voor Human Factors*, 41(1):9–14, 2016.
- [Tuf01] Edward R. Tufte. *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, CT, 2001.
- [Tuk77] John W. Tukey. *Exploratory data analysis*. Addison-Wesley, Reading, MA, 1977.

- [Ula00] Robert E. Ulanowicz. *Handbook of Ecosystem Theories and Management*, chapter Ascendancy: A Measure of Ecosystem Performance, pages 303–315. Lewis Publishers, Boca Raton, FL, first edition, 2000.
- [VL21] Giulia Vilone and Luca Longo. Notions of explainability and evaluation approaches for explainable artificial intelligence. *Information Fusion*, 76:89–106, December 2021.
- [vW05] J.J. van Wijk. The value of visualization. In *IEEE Visualization (VIS)*, pages 79–86, Minneapolis, MN, October 2005.
- [WAF16] Ashley Wheat, Simon Attfield, and Robert Fields. Analytic provenance and distributed sensemaking. In *Building Trust in Information*, Springer Proceedings in Business and Economics (SPBE), pages 151–161, August 2016.
- [WAM⁺19] Emily Wall, Meeshu Agnihotri, Laura Matzen, Kristin Divis, Michael Haass, Alex Endert, and John Stasko. A heuristic approach to value-driven evaluation of visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):491–500, January 2019.
- [WC02] Mark A. Whiting and Nick Cramer. WebthemeTM: Understanding web information through visual analytics. In *The Semantic Web — ISWC*, pages 460–468, Sardinia, Italy, January 2002.
- [WCR20] Dale F. Whelehan, Kevin C. Conlon, and Paul F. Ridgway. Medicine and heuristics: Cognitive biases and medical decision-making. *Irish Journal of Medical Science (1971 -)*, 189:1477–1484, May 2020.
- [WDC⁺23] Aoyu Wu, Dazhen Deng, Min Chen, Shixia Liu, Daniel Keim, Ross Maciejewski, Silvia Miksch, Hendrik Strobel, Fernanda Viégas, and Martin Wattenberg. Grand challenges in visual analytics applications. *IEEE Computer Graphics and Applications*, 43(5):83–90, Sep.-Oct. 2023.
- [Wea64] Warren Weaver. Scientific explanation. *Science*, 143(3612):1297–1300, March 1964.
- [Wei22] Daniel Weiskopf. Uncertainty visualization: Concepts, methods, and applications in biological data visualization. *Frontiers in Bioinformatics*, 2, February 2022.
- [Wel76] Roy E. Welsch. Graphics for data analysis. *Computer & Graphics*, 2(1):31–37, 1976.
- [Wel19] Adrian Weller. *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, chapter Transparency: Motivations and Challenges, pages 23–40. Springer, 2019.

- [WGK10] Matthew Ward, Georges Grinstein, and Daniel Keim. *Interactive Data Visualization: Foundations, Techniques and Applications*. A.K. Peters, Natick, MA, 2010.
- [WHL⁺19] Andrew Wentzel, Peter Hanula, Timothy Luciani, Baher Elgohari, Hesham Elhalawani, Guadalupe Canahuate, D Vock, Clifton D Fuller, and G Elisabeta Marai. Cohort-based T-SSIM visual computing for radiation therapy prediction and exploration. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):949–959, January 2019.
- [WHLS19] Junpeng Wang, Subhashis Hazarika, Cheng Li, and Han-Wei Shen. Visualization and visual analysis of ensemble data: A survey. *IEEE Transactions on Visualization and Computer Graphics*, 25(9):2853–2872, September 2019.
- [WJBC19] Marcin Wójcik, Pamela Jeziorska-Biel, and Konrad Czapiewski. Between words: A generational discussion about farming knowledge sources. *Journal of Rural Studies*, 67:130–141, April 2019.
- [WNC⁺22] Emily Wall, Arpit Narechania, Adam Coscia, Jamal Paden, and Alex Endert. Left, right, and gender: Exploring interaction traces to mitigate human biases. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):966–975, January 2022.
- [WP91] Sandra S. Walther and Richard L. Peskin. Object-oriented data management for interactive visual analysis of three-dimensional fluid-flow models. In *Extracting Meaning from Complex Data: Processing, Display, Interaction II*, volume 1459, pages 232–243, San Jose, CA, 1991.
- [WSL⁺20] Andreas Walch, Michael Schwärzler, Christian Luksch, Elmar Eisemann, and Theresia Gschwandtner. LightGuider: Guiding interactive lighting design using suggestions, provenance, and quality visualization. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):569–578, January 2020.
- [WT04] Pak Chung Wong and Jim Thomas. Visual analytics. *IEEE Computer Graphics and Applications*, 24(5):20–21, Sep.-Oct. 2004.
- [WWHC11] Ralph Weissleder, Jack Wittenberg, Mukesh G. Harisinghani, and John W. Chen, editors. *Primer of Diagnostic Imaging*, chapter Chapter 14: Imaging Physics, pages 690–746. Mosby, Philadelphia, PA, sixth edition, 2011.
- [WXK22] Emily Wall, Cindy Xiong, and Yea-Seul Kim. VisHikers’ guide to evaluation: Competing considerations in study design. *IEEE Computer Graphics and Applications*, 42(3):29–38, May-Jun. 2022.

- [XAJK⁺15] Kai Xu, Simon Attfield, T.J. Jankun-Kelly, Ashley Wheat, Phong H. Nguyen, and Nallini Selvaraj. Analytic provenance for sensemaking: A research agenda. *IEEE Computer Graphics and Applications*, 35(3):56–64, May-Jun. 2015.
- [XOW⁺20] Kai Xu, Alvitta Ottley, Conny Walchshofer, Marc Streit, Remco Chang, and John Wenskovitch. Survey on the analysis of user interactions and visualization provenance. *Computer Graphics Forum*, 39(3):757–783, June 2020.
- [Xu11] Junming Xu. *Topological Structure and Analysis of Interconnection Networks*. Springer, New York, NY, 2011.
- [YCON25] Fernando Yanez, Cristina Conati, Alvitta Ottley, and Carolina Nobre. The state of the art in user-adaptive visualizations. *Computer Graphics Forum*, 44(1):e15271, February 2025.
- [YDH19] Fan Yang, Mengnan Du, and Xia Hu. Evaluating explanation without ground truth in interpretable machine learning. arXiv, 2019.
- [YJ95] L. Richard Ye and Paul E. Johnson. The impact of explanation facilities on user acceptance of expert systems advice. *MIS Quarterly*, 19(2):157–172, June 1995.
- [You32] Kimball Young. General and specific attitudes. hadley cantril. *American Journal of Sociology*, 38(2):282–285, September 1932.
- [YZE⁺24] Lixiang Yan, Linxuan Zhao, Vanessa Echeverria, Yueqiao Jin, Riordan Alfredo, Xinyu Li, Dragan Gašević, and Roberto Martinez-Maldonado. VizChat: Enhancing learning analytics dashboards with contextualised explanations using multimodal generative AI chatbots. In *Artificial Intelligence in Education*, volume 14830 of *Lecture Notes in Computer Science (LNCS)*, pages 180–193, July 2024.
- [ZAW⁺18] Jianlong Zhou, Syed Z. Arshad, Xiuying Wang, Zhidong Li, Dagan Feng, and Fang Chen. End-user development for interactive data analytics: Uncertainty, correlation and user confidence. *IEEE Transactions on Affective Computing*, 9(3):383–395, Jul.-Sep. 2018.
- [ZKL17] Maryam Zaghian, Azin Khabazian, and Gino Lim. A chance-constrained programming framework to handle uncertainties in radiation therapy treatment planning. *European Journal of Operational Research*, 266(2):736–745, October 2017.
- [ZPF23] Abdullah Kaan Zaimoglu, Lorien Pratt, and Brian Fisher. Epistemological role of human reasoning in data-informed decision-making. *Frontiers in Communication*, 8, November 2023.

Curriculum Vitae

Maath Musleh Musleh

maath@cg.tuwien.ac.at

Beit Safafa, Jerusalem, Palestine

Summary/Profile

Maath Musleh is a researcher with a strong interdisciplinary interest and experience. He has an educational background in Computer Science and Political Journalism. He also has an academic working experience in Humanities and Arts and a research experience in the Social Sciences. His current research interest spans around user behavior in visualization and visual analytics. In his PhD, he researches user attitude dynamics within decision making in visual analytics under uncertainty. He looks into methods and techniques to guide the user in the sense-making while efficiently building user trust and support his decision confidence. His interdisciplinary interest and research are complemented by strong technical skills, especially in Python.

Education

TU WIEN

[ONGOING] PhD in Visualization and Visual Analytics

[Thesis] Guided Visual Analytics for Decision Making under Uncertainty

Vienna, Austria

2021 – 2025

LINNAEUS UNIVERSITY

MSc in Media Technology, Social Media and Web Technologies

[Thesis] Visual Analysis of Industrial Multivariate Time-Series Data

Holder of the Swedish Institute Scholarship for Global Professionals (SISGP)

Växjö, Sweden

June 2021

THE CITY UNIVERSITY

MA in Political Journalism

London, England

October 2012

THE AMERICAN UNIVERSITY IN DUBAI

Bachelor in Information Technology

Concentration: Application Development for Commerce and Industry

Dubai, UAE

May 2007

Research Interests

- Decision making in Visual Analytics
- Uncertainty in Visual Analytics
- User Attitude (Confidence and Trust) in Visual Analytics
- Visual Analytics in Agriculture and Farming
- Other interdisciplinary fields in Visualization and Visual Analytics

Publications, Reviews, Conferences

- **Journal and Conference papers and posters [8]:**
 - **Maath Musleh**, Renata G. Raidou, Davide Ceneda, “**TrustME: A Context-Aware Explainability Model to Promote User Trust in Guidance**”, in IEEE Transactions on Visualization and Computer Graphics, Volume 31, No.10, 2025, pp. 8040-8056, doi:[10.1109/TVCG.2025.3562929](https://doi.org/10.1109/TVCG.2025.3562929). [Invited Talk: IEEE Vis 2025]
 - **Maath Musleh**, Davide Ceneda, Henry Ehlers, Renata G. Raidou, “**ConAn: Measuring and Evaluating User Confidence in Visual Data Analysis Under Uncertainty**”, in Computer Graphics Forum, Volume 44, No.1, 2025, e15272, doi:[10.1111/cgf.15272](https://doi.org/10.1111/cgf.15272). [Invited Talk: EuroVis 2025]
 - **Maath Musleh**, Ludvig Paul Muren, Laura Toussaint, Anne Vestergaard, Eduard Gröller, Renata G. Raidou, “**Uncertainty guidance in proton therapy planning visualization**”, in Computers & Graphics, Volume 111, 2023, pp. 166-179, doi:[10.1016/j.cag.2023.02.002](https://doi.org/10.1016/j.cag.2023.02.002). [Invited Talk: EG VCBM 2023]
 - **Maath Musleh**, Angelos Chatzimparmpas, Ilir Jusufi, “**Visual analysis of blow molding machine multivariate time series data**”, in Journal of Visualization, Volume 25, 2022, pp. 1329–1342, doi:[10.1007/s12650-022-00857-4](https://doi.org/10.1007/s12650-022-00857-4).
 - **Maath Musleh**, Angelos Chatzimparmpas, Ilir Jusufi, “**Visual Analysis of Industrial Multivariate Time Series**”, in the proceedings of *the 14th International Symposium on Visual Information Communication and Interaction (VINCI '21)*, Potsdam, Germany, pp. 1–5, doi:[10.1145/3481549.3481557](https://doi.org/10.1145/3481549.3481557). [Best Short Paper Award] [Conference Talk]
 - **Francesca Gauci** and **Maath Musleh**, “**Agritology: A Decision Support System for Local Farmers in Malta and Palestine**”, presented at the 3rd Linnaeus Student Conference on Information Technology (LSCIT), Växjö, Sweden, pp. 1–13. [Conference Talk]
 - **Francesca Gauci** and **Maath Musleh**, “**Agritology: A Decision Support System for Local Farmers in Malta and Palestine**”, presented at the 3rd International Workshop on Semantics for Biodiversity (S4BioDiv), Bolzano, Italy, pp. 1–5, <http://ceur-ws.org/Vol-2969/paper3-s4biodiv.pdf>. [Poster]
 - **Maath Musleh**, “**The Dynamics of Peace and Justice in Palestine,**” presented at Trinity College Dublin: Peacebuilding & Local Communities in Post-Conflict Societies, May 7, 2019, Dublin, Ireland, doi:[10.13140/RG.2.2.30120.69122](https://doi.org/10.13140/RG.2.2.30120.69122)

- **Conference Talks:**
 - Bard College: Global | Local: Experiments in the Arts and Humanities 2017, New York [**Research Concept:** Challenges & experiments in developing and co-teaching the “Thinking about Video Games” course]
 - Bard College: Digital & Visual Culture in the Liberal Arts 2015, Istanbul [**Research Concept:** Palestinian Visual Narrative & the Formation of a Collective Memory]
- **Books:** [Strategic Studies] The Zionist Policies to Combat ‘BDS’ and the Ways to Dismantle Them (in Arabic), 2018, 978-9950-400-13-9, MASARAT
- **Paper Reviews [14]:** IEEE VIS 2023-2025, CESC 2024-2025, EuroVis 2024-2025, PacificVis 2023-2024, ICSTCC 2023, CESC 2024-2025, ICSTCC 2023, Cognition, Technology & Work (CTM) Journal.
- **Posters Review Committee [5]:** IEEE Vis 2025.
- **Committees:** IPC for the IEEE VIS poster session 2025.

Academic Experience

TECHNICAL UNIVERSITY OF VIENNA, TU WIEN

Vienna, Austria

University Assistant

Nov 21 – Present

- Participate in research tasks, in teaching and administrative tasks
- Teach and conduct lectures
- Participate in examinations
- Teaching Assistant and Lecturer for the following courses:
 - Visualization 1 (Lab Coordinator) (2023W)
 - Visualization 2 (Lecture: Uncertainty Visualization) (2023W, 2024W)
 - Methods for Data Generation and Analytics (Lab Coordinator) (2023W, 2024W)
 - CG/Visualization Seminar (Students Project Supervisor) (2022S/W, 2023S/W, 2024S/W, 2025S)
 - Information Visualization (Lab Coordinator: 2022S, 2023S, 2024S, 2025S)

THE (OSUN) HUBS FOR CONNECTED LEARNING INITIATIVES

Jordan

Academic Consultant

Aug 2021 – May 2022

- Participating in planning meetings with local universities and implementing partners in the MENA region;
- The development and delivery of three online courses, built on the OSUN-appointed platform and replicable in the future (1. Social Media Ecology, 2. TBD, 3. TBD);
- Liaising with Jordan-based partners and the OSUN Hubs Program Assistant to support Jordan-based Hub students;
- Working directly with university partners and the Hubs Leadership to establish BA and diploma pathways for learners.

CRITICAL PRACTICE STUDIO IN DHEISHEH ON “TRANSLATION” Palestine
Academic and Creative Facilitator Jul 13 – 19, 2021

- Facilitate workshops
- Provide academic and creative support to the residents

PARAMI UNIVERSITY ONLINE (Myanmar)

Instructor

Writing and Thinking Intensive Course (30 hours) Jan 2021

- Teach one sections that consist of 17 students

OSUN MICROCOLLEGES FOR CONNECTED LEARNING INITIATIVE ONLINE

Instructor

Language and Thinking Intensive Course Aug 2020 – Sep 2020

- Teach two sections that consist of 10-16 students each from different parts of Kenya and Jordan

ALQUDS BARD COLLEGE FOR ARTS & SCIENCES Jerusalem, Palestine

Visiting Instructor

Humanities and Arts Division Jan 2019 – Aug 2019

- Teach - Media Law & Ethics Course (Spring '19) and Literature of Exile & Migration (Summer '19)
- Participate in faculty and curriculum development meetings and workshops
- Co-advise Senior projects in the Humanities and Arts & the Social Sciences Divisions

ALQUDS BARD COLLEGE FOR ARTS & SCIENCES Jerusalem, Palestine

Head of the Humanities and Arts Division

Head of the Media Studies Program Aug 2016 – Jul 2017

- Developed the Media Studies Program plan
- Headed the division that encompassed two programs, 6 faculty members, +30 students

Coordinator of the Arabic Language and Thinking Program May 2016 – Jun 2017

- Administered the teaching of the 2-weeks intensive core course to 672 freshmen students at Al-Quds University
- Compiled the Anthology for the course
- Overlooked the training of 45 new faculty members
- Developed the academic structure of the program
- Supervised the class teaching plans of 45 sections

Acting Head of the Media Studies Program Aug 2015 – Jul 2016

Acting Head of the First-Year Seminar Program (College Core) Aug 2015 – Dec 2015

Visiting Lecturer in the College Core Program Jan 2014 – Jul 2017

Visiting Lecturer in the Media Studies Program

Sep 2012 – Jul 2017

- Taught several core, elective, and cross-listed courses: *Introduction to Media Studies, Media Law & Ethics, New Media, Topics in the Business of Media, First Year Seminar I & II, HR & the Media, Electronic Discourses, Thinking about Video Games, Understanding Media Art (Intensive English students), Video I, Advanced Digital Editing*
- Advised +50 students on their academic development and registration
- Advised several production-based senior projects in the Media Studies program and senior theses in the Social Sciences and the Humanities divisions
- Served on the senior boards of Computer Science, Biology, Economics, Political Science, Human Rights, Urban Studies, Literature, and Media Studies students

Trainings and Workshops

BUCHAREST UNIVERSITY OF ECONOMIC STUDIES

Predeal, Romania

Data Science and AI International Summer School 2025

Jul 2025

TU WIEN (INTERNAL LEARNING)

Vienna, Austria

How do you write a data management plan (DMP) (3hrs)

May 2025

Information Security

Jan 2025

Sexual Harassment at University and at Work (2hrs)

Jan 2025

Unconscious Bias for Staff (2hrs)

Jan 2025

Diversity Basics for Staff (2hrs)

Jan 2025

ORCID and your digital identity as a researcher (1hr)

Jan 2025

Core Competency AI – Promoting Students' AI and Media Skills (4hrs)

Jan 2025

Improving Your Scientific Papers (8hrs)

Nov 2024

First Aid (8hrs)

Sep 2024

Fit for AI: Future Skills in Research & Education (4hrs)

Jun 2024

Didactic Basics (16hrs)

Sep 2023

WELFARE ASSOCIATION (TAAWON)

Jerusalem, Palestine

Introduction to the Historical and Architectural Development of Palestine

Apr 2025

EG VCBM 2023

Norrköping, Sweden

Bio+Med+Vis Summer School

Sep 2023

BARD INSTITUTE FOR WRITING & THINKING

Teaching Pedagogy Workshop, participant

Abu Dis, Palestine - Jan 2020

The Reluctant Learner Workshop, participant

Abu Dis, Palestine - Aug 2017

Applying the Practices workshop, participant

Annandale-on-Hudson, NY - Jul 2016

Language and Thinking workshop, participant

Annandale-on-Hudson, NY - Jul 2014

HUMANITY CREW Psychosocial support for refugees	Haifa, Palestine Dec 2018
AL-QUDS BARD COLLEGE FOR ARTS AND SCIENCES Writing to Learn workshop, participant	Jerusalem, Palestine Jan 2014, Jan 2015
GLOBAL VOICES The 4 th Arab Bloggers Meeting, Participant	Amman, Jordan Jan 2014
HEINRICH BOLL Energize, Polarize, Mobilize! International Conference, Participant	Berlin, Germany Mar 2013
PALVISION and THE BRITISH CONSULATE Active Citizens Programme, Participant	Ramallah, Palestine Dec 2010
YOUNG ARAB LEADERS Arab Leadership Conference, Participant	Beirut, Lebanon May 2009
Other Professional Experience	
THE PALESTINE FESTIVAL OF LITERATURE Logistics Manager - Contract	Jordan/Palestine Dec 2018 – May 2019
FREELANCER Journalist, Producer & Fixer	Palestine Oct 2010 – Jun 2014
TRT TV Producer & Reporter	Jerusalem, Palestine Oct 2013 – Dec 2013
4D MEDIA PRODUCTION AND SERVICES Co-founder	Ramallah, Palestine Jun 2010 – May 2011
Volunteering and Activities	
TU WIEN VCBM 2022 Conference	Vienna, Austria Sep 2022
OM SLEIMAN ORGANIC FARM Trainee Volunteer (Farming)	Bilin, Palestine May – Jul 2018

YABOUS CULTURAL CENTER Volunteer (Leader – Cultural Events)	Jerusalem, Palestine Oct 2018
RELIEF AND RECONCILIATION Volunteer (Refugee aid)	Akkar, Lebanon Aug 2017 – Oct 2017
PALESTINE FESTIVAL OF LITERATURE Palestine Social Media Coordinator, Advisor	May 2013 – May 2017
MOSSAWA CENTER (International Volunteer Summer Camp) Volunteer, Leader (Social work)	Palestine 2011 – 2013
THE AMERICAN UNIVERSITY IN DUBAI Student Government Association, Sports Events Coordinator Student Ambassador, Orientation Leader, Supervisor	Dubai, U.A.E. 2004 – 2005 2004 – 2007

Skills

Programming:

Proficient: Python (experience with Kivy, Django, Scikit-learn, Dash and Plotly Libraries), R (experience with statenet and igraph libraries), HTML5/CSS

Knowledge and background: C#, PHP, JavaScript (D3), SQL, ASP.net, AL, C/AL

Software:

Adobe Premiere, Audition, Photoshop, Final Cut Pro, ERP System: Microsoft Dynamic 365

Languages: Arabic (Native), English (Fluent), French (A2 - Basic), German (A2 - Basic)

Hobbies and Interests: Farming, Football, Basketball, Hiking and Camping, Fishing, Reading (especially, History and Philosophy), Socio-political causes, Table Tennis, Cooking & Baking.

References can be acquired by contacting:

- Renata Raidou (rraidou@cg.tuwien.ac.at)
- Davide Ceneda (d.ceneda@tue.nl)
- Eduard Gröller (groeller@cg.tuwien.ac.at)