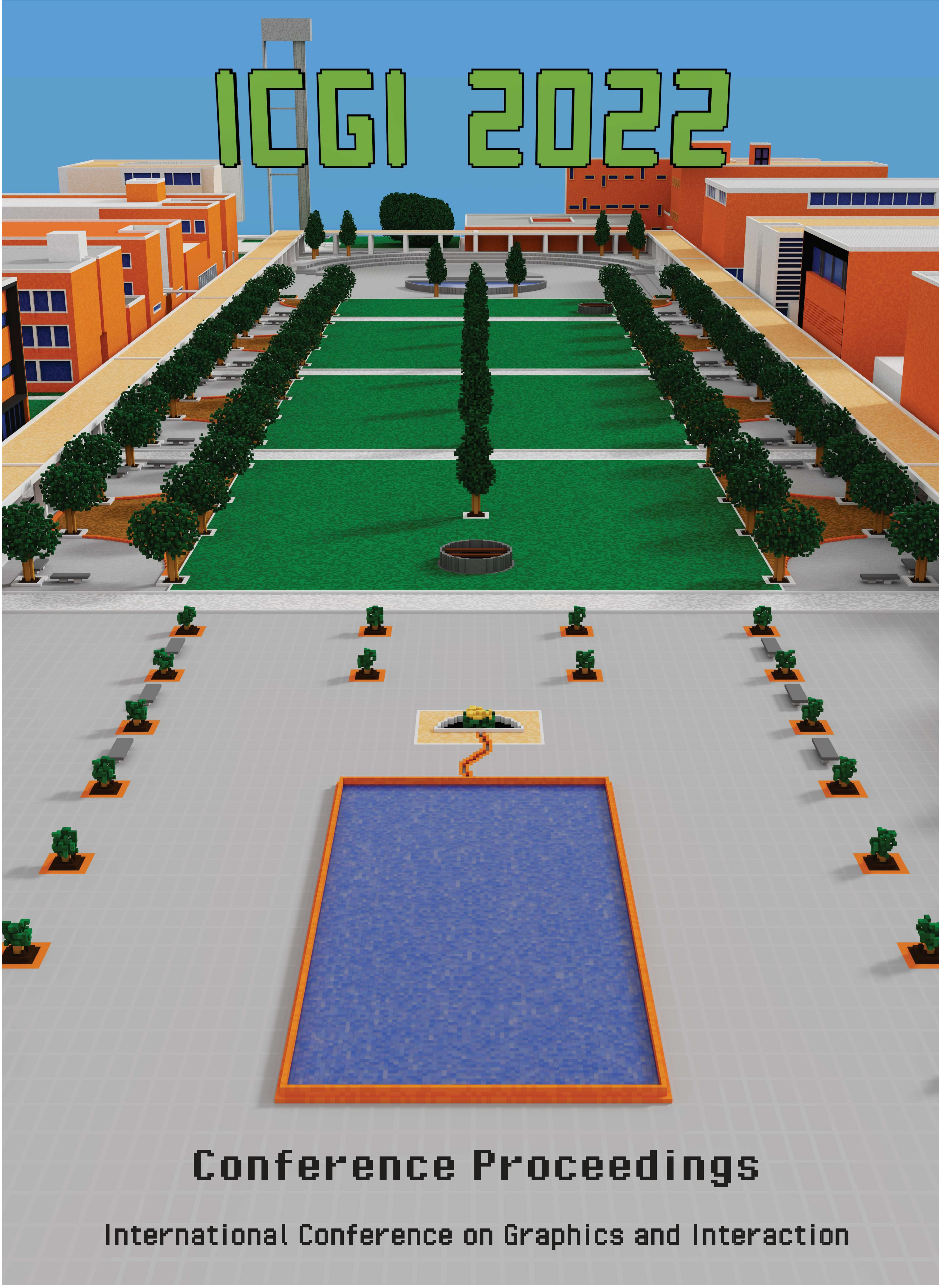


ICGI 2022

Conference Proceedings

International Conference on Graphics and Interaction

universidade de aveiro

deti departamento de eletrónica, telecomunicações e informática

ieeta

deca departamento de comunicação e arte

DigiMedia digital media and interaction research centre

Grupo Português de Computação Gráfica Eurographics Portuguese Chapter

Sponsored by: BOSCH MINDERA GAMING

Proceedings of the

ICGI 2022 - 4th International Conference on Graphics and Interaction

November 3-4, 2022

University of Aveiro, Aveiro, Portugal

Proceedings of the ICGI 2022 - 4th International Conference on Graphics and Interaction

Luís Gonzaga Magalhães and Paula Alexandra Silva (editors)

979-8-3503-4573-5

November 2022

LaTeX class and proceedings files based on Giampiero Salvi's
(https://github.com/giampierosalvi/LaTeXProceedings)

Preface

It is with great pleasure that we welcome you to the 2022 edition of the International Conference on Graphics and Interaction, held on November 3-4, 2022, at the University of Aveiro, Portugal, as a joint organization with the Eurographics Portuguese Chapter-GPCG.

ICGI' 2022 aims to bring together researchers, teachers, and professionals in the areas of Computer Graphics, Image Processing, Computer Vision and Human-Computer Interaction, allowing the dissemination of concluded or ongoing work, as well as the exchange of experiences between the academic, industrial, and end-user communities.

Similar to last year, this event includes a Computers & Graphics journal special section on Recent Advances in Graphics and Interaction. From the 9 submissions to this special section, 3 were accepted to the Computers & Graphics journal.

The ICGI' 2022 had 31 conference papers (28 long, 3 short) and 3 poster submissions. From the conference papers, 12 were accepted as long papers, 9 as short papers, and 6 invited as posters, following a double-blind review process. These 21 contributions (12 long, 9 short) will be presented at the conference, as well as the 3 accepted posters. Similar to previous years, long papers will be indexed and available at the IEEE Xplore Digital Library.

It is also with great pleasure that we thank the presence of the invited keynote speakers, Sergi Bermúdez I Badia, from the University of Madeira, Portugal, and Abel Gomes, from the University of Beira Interior, Portugal. Our most sincere thanks for accepting our invitation and enriching this event.

We would also like to thank all those who contributed to this event, including the authors, scientific committee members, student volunteers, institutional organizations, and sponsors. We wish you a very productive and exciting event!

Kind regards,

Luís Gonzaga Magalhães (UM, Algoritmi)
Paula Alexandra Silva (UC - DEI, CISUC)
(ICGI 2022 Program Committee chairs)

Aveiro, November 2022.

Scientific Committee

Program Chairs

Luís Gonzaga Magalhães (UM, algoritmi)
Paula Alexandra Silva (UC, dei)

Members

A. Augusto Sousa (University of Porto)
Abel Gomes (University of Beira Interior, Department of Computer Science and Engineering)
Adérito Fernandes-Marcos (University of Saint Joseph, Macao)
Alan Chalmers (The University of Warwick)
Alexandrino Gonçalves (ESTG Leiria)
Alfredo Ferreira (IST/INESC-ID/ULisboa)
Ana Cláudio (LASIGE, Faculdade de Ciências, Universidade de Lisboa)
Ana Paula Afonso (LASIGE, Faculdade de Ciências, Universidade de Lisboa)
Anabela Marto (University of Porto)
André Rangel (CITAR-UCP-FBAUP)
Andrea Giachetti (University of Verona)
António Coelho (University of Porto)
António Ramires Fernandes (University of Minho)
Augusto Esteves (ITI / LARSyS, Instituto Superior Técnico, University of Lisbon)
Beatriz Sousa Santos (University of Aveiro)
Bernardo Marques (University of Aveiro)
Brígida Mónica Faria (Higher School of Health/Polytechnic of Porto (ESS/P.Porto))
Carlo Harvey (Birmingham City University)
Daniel Gonçalves (INESC-ID)
Daniel Mendes (INESC TEC, FEUP)
David Narciso (INESC TEC)
David Lamas (Tallinn University)
Diogo Cabral (ITI/LARSyS, IST, University of Lisbon)
Duarte Duque (2Ai, Polytechnic Institute of Cavado and Ave)
Fernando Birra (Universidade NOVA de Lisboa)
Frutuoso Silva (University of Beira Interior)
Guilherme Gonçalves (INESC TEC)
Hugo Coelho (University of Trás-os-Montes e Alto Douro)
João Brisson (Universidade de Lisboa)
João Jacob (University of Porto)
João Madeiras Pereira (IST/INESC-ID)
João Moreira (Universidade de Lisboa)
Joaquim Madeira (University of Aveiro)
José Braz Pereira (EST Setúbal / IPS)
José Creissac Campos (University of Minho & HASLab/INESC TEC)
José Luís Silva (ISCTE-IUL, University of Madeira / Madeira-ITI)
Kadi Bouatouch (IRISA)
Leonel Morgado (INESC TEC / Universidade Aberta)
Lidia Ortega (University of Jaén)
Luis Santos (University of Minho)
Luis Romero (Instituto Politécnico de Viana do Castelo)
Luis Gonzaga Magalhaes (University of Minho)

Manuel J. Fonseca (Universidade de Lisboa)
Maria Beatriz Carmo (Universidade de Lisboa)
Mauricio Sousa (University of Toronto)
Miguel Angel Guevara Lopez (EST Setúbal / IPS)
Miguel Carvalhais (INESC TEC / Faculdade de Belas Artes, Universidade do Porto)
Miguel Melo (Universidade de Trás-os-Montes e Alto Douro)
Miguel Sales Dias (ISCTE)
Nelson Zagalo (University of Aveiro)
Nuno Rodrigues (ESTG-Politécnico de Leiria)
Óscar Mealha (University of Aveiro)
Paula Alexandra Silva (University of Coimbra)
Paulo Dias (University of Aveiro)
Pedro Amado (Fine Arts Faculty, University of Porto)
Pedro Miguel Moreira (ESTG – IPVC & LIACC)
Ricardo Marques (Universitat de Barcelona)
Roberto Lam (ISE – Universidade do Algarve)
Rui Nóbrega (Universidade NOVA de Lisboa)
Rui Rodrigues (University of Porto)
Samuel Silva (University of Aveiro)
Sandra Gama (Instituto Superior de Tecnologias Avançadas de Lisboa)
Soraia Paulo (IST/UL/INESC-ID)
Tânia Rocha (University of Trás-os-Montes and Alto Douro)
Telmo Adão (Centro de Computação Gráfica)
Teresa Chambel (Universidade de Lisboa)
Teresa Romão (Faculdade de Ciências e Tecnologia/Universidade NOVA de Lisboa (FCT/UNL))
Thomas Bashford-Rogers (The University of Warwick)
Tomás Alves (Universidade de Lisboa)
Vítor Sá (Universidade Católica Portuguesa)

Organizing Committee

Conference Chairs

Paulo Dias (UA, deti)
Joaquim Madeira (UA, deti)
Nelson Zagalo (UA, deca)
Mário Vairinhos (UA, deca)

Web Chairs

Bernardo Marques (UA, DigiMedia/IEETA)
Pedro Reisinho (UA, deca)

Organization, Support and Sponsors

Organizers

Support

IEEE - Institute of Electrical and Electronics Engineers

Sponsors

Contents

Keynote Speakers

Sergi Bermúdez I Badia

University of Madeira

Currently, he is an Associate Professor (tenure) at the University of Madeira, where he teaches for the Informatics and Interactive Media Design Masters; researcher of the NOVA Laboratory for Computer Science and Informatics and coordinator of its N-LINCS branch in Madeira, and president of the International Society for Virtual Rehabilitation. He received his Msc. in telecommunications engineering from the Universitat Politecnica de Catalunya (UPC) and a PhD from the Swiss Federal Institute of Technology Zürich (ETHZ).

He has pursued research at several institutes in Europe and the USA, including the Laboratoire de Production Microtechnique at the EPFL (Lausanne), the Institute of Neuroinformatics at the ETHZ (Zurich), at the Institute of Audiovisual Studies at the Technology Department of the Universitat Pompeu Fabra (Barcelona), where he was a Juan de la Cierva research fellow and head of the Robotic Systems Laboratory at the laboratory for Synthetic Perceptive, Emotive and Cognitive Systems (SPECS), and the Quality of Life Technologies and Entertainment technology centers of the Carnegie Mellon University (Pittsburgh).

His scientific goal is to investigate biological systems' underlying neural mechanisms and exploit them using real-world artefacts, with particular emphasis on neuro-rehabilitation systems, interactive technologies, and robots.

Web page: https://www.uma.pt/directorio/sergi-bermudez-i-badia/

Personalized Applications of VR to Stroke Rehabilitation and Fitness Training

Nowadays, it is widely accepted that games, and entertainment technologies in general, have very interesting features that, if used properly, can largely contribute to the effectiveness of treatments in different health domains. These games, also known as games-with-a-purpose, need to achieve a very difficult and interesting balance among science, health, engineering and entertainment. In this talk, I will present the approach we follow at the NeuroRehabLab, where we combine games, Human-Computer Interaction and clinical rehabilitation guidelines to develop interactive systems that are novel and effective tools for motor and cognitive rehabilitation, with special emphasis on stroke. I will discuss the effect of interface technology on motor-cognitive interference in task performance; a participatory design approach with health professionals to develop parameterized models for the training of Activities of Daily Living in a simulated environment; and how we automate the parameter selection process in these games by means of an adaptive approach. This strategy allows these systems to be used by patients with different cognitive and motor skills while still providing personalized training.

Abel Gomes

University of Beira Interior

Abel Gomes is an Associate Professor in Computer Graphics and Games at the Department of Computer Science, University of Beira Interior, Portugal, and a senior researcher at INESC-ID (Graphics and Interaction Group), Lisbon, Portugal. He is also Associate Editor of Computers & Graphics (Elsevier) and International Journal of Computer Games Technology (Hindawi-Wiley).

His research areas of interest include Geometric Modeling, Computational Geometry, Computer Graphics, Computer Games, Computational Biology, and Molecular Graphics, and Augmented Reality in Medicine. During his career, he has approached a number of scientific problems, namely convex hull algorithms, pathfinding algorithms, linearization of implicit curves, triangulation of implicit surfaces, triangulation of molecular surfaces, protein pocket detection methods, and quantum molecular dynamics. Currently, his main research topic is surface reconstruction using both classical and deep learning algorithms.

Web page : http://www.di.ubi.pt/~agomes/

Surface Reconstruction from Point Clouds: Past, Present, and Future

Reconstructing object/scene surfaces is an important, long-standing problem in computer vision and graphics research. Its applications range from computer-aided design, computer animation, robotics, virtual/augmented reality, or even medical engineering. When dealing with this problem, we assume that the surface to be reconstructed from a point cloud is a 2D manifold embedded in the 3D Euclidean space. However, this problem is ill-posed because there are infinitely many solutions or surfaces for the same point cloud. Even worse is the fact that many point clouds may present noise, non-uniform point distribution, missing point chunks, outliers, and the like, as a result of some scanning issues during the data acquisition process.

Consequently, and considering both classical and deep learning algorithms, there is no reconstruction surface algorithm, capable of correctly reconstructing a surface from any 3D scanned point cloud. Surprisingly, or maybe not, a few classical methods perform even better than deep learning methods in terms of both robustness and generality. In a way, this talk aims to overview the state-of-the-art in surface reconstruction, hoping to open some research avenues for the future.

Session 1: Visualization and Interaction

chair: Nelson Zagalo

Examining User Preferences based on Personality Factors in Graphical User Interface Design

Tomás Alves, Daniel Gonçalves, Sandra Gama
HUMAN Lab, Graphics and Interaction, INESC-ID, IST,
University of Lisbon, Portugal
{tomas.alves,daniel.j.goncalves,sandra.gama}@tecnico.ulisboa.pt

Joana Henriques-Calado
CICPSI, Faculdade de Psicologia,
University of Lisbon, Portugal
joanahenriquescalado@gmail.com

***Abstract*—Individual differences play a major role in human-computer interaction. In particular, personality shapes how we process and act on the world, and how users perceive and accept technology. Nevertheless, there is limited evidence on the effect of different personality types in graphical user interface design preferences. Weighting how personality affects perception, we leverage in-depth synergies between personality variables and design preferences for graphical user interface elements to study whether it is possible to formulate a novel set of design guidelines that allow the creation of user interfaces customized to psychological variables. A clustering approach of the subjects (N=65) yielded three different personality profiles based on the personality variables of the Five-Factor Model. Then, an association rules algorithm produced a set of rules from which we created a set of design guidelines. We discuss the study implications and future work opportunities.**

***Index Terms*—personality, user preferences, design guidelines, user study, graphical user interfaces**

I. Introduction

Psychology principles have been notably applied as a core piece of Human-Computer Interaction (HCI) research. In particular, recent work has focused on informing design choices and understand differences regarding how individuals use technology (e.g., [1]). It enables researchers to take conclusions regarding design effectiveness, since successful technology development needs input from a representative set of potential users and, more precisely, the range of differences among individuals may influence technology [2]. Some factors may include age, gender, job function, language culture or fundamental idiosyncratic attributes, such as personality and motivation. The inclusion of these factors empowers developers to take into account not only how individual characteristics of the user impact the user experience (UX), but also consumers' expectations from providers across a large range of fields. However, there is limited evidence of the usefulness of designing a graphical user interface (GUI) based on individual psychological variables [3].

Weighting how personality affects perception [4] and design efficiency [5], we focus on how in-depth synergies between personality variables and graphical user interface preferences can be applied to graphical user interface design to accommodate the preferences of diverse users. The potential of GUI design based on personal characteristics has been studied by customizing the display to meet certain demands [6]. Sarsam and Al-Samarraie [7] focused on how differences in personality traits can stimulate individuals' information processing capabilities according to their display preferences. The authors focused on the Five-Factor Model (FFM) [8], which categorizes personality with five traits: neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness. In particular, the authors grouped individuals with common personality profiles in two clusters – one addressing neuroticism, and the other extraversion and conscientiousness – and continued with an association rules technique to find which design elements were preferred for each group of subjects. Results showed how the visual experience improves when subjects interacted with the interface designed based on their personality characteristics. However, there is no set of complete GUI design guidelines to explain the preferences regarding certain interface design features. Additionally, only a small number of interface elements that cover a limited variation of styles have been studied.

In this light, our research goal is **to create a set of design guidelines based on a match between user preferences for GUI feature styles and personality factors.** In particular, we extend prior state-of-the-art research of Alves et al. [3] and Sarsam and Al-Samarraie [9] by studying in-depth personality variables from the FFM, since other current research only applies personality at a superficial trait level, thus neglecting facets, a specific and unique aspect of a broader personality trait, that may provide far more detailed insights into the relationship we are addressing. Our contribution adds new key pieces of knowledge to the field of HCI through the applied methodology. To the best of our knowledge, our work is the first to replicate the personality profiling and GUIs creation based on Sarsam and Al-Samarraie [9] in the website layout context. Although we do not intend to validate personality profiling, our results provide more insights regarding how this methodology is relevant for the introduction of personality in the design process.

II. Related Work

Recent research leverages individual characteristics to personalize user interface (UI)s and improve user experience [7], [10]. In particular, GUIs designed in accordance to user personality have been shown to affect both information-seeking performance and behavior [11], [12], as well as user

978-8-3503-4573-57/22/$31.00 ©2022 IEEE

preference [13]. There is a wide variety of graphical elements that can be customized such as structure [14], navigation [15], layout [16], font style attributes [17], font size [15], [18], buttons [19], color [20], [21], list [22], information density [20], support [20], and alignment [23]. As such, it is of utmost importance to consider in-depth how the designers should draft the GUI, since personality factors affect information-seeking behaviors and, in particular, how one builds their mental model to interact with a piece of technology [11], [12]. However, we believe that there is limited evidence of the effect of designing GUIs based on individual psychological variables [3]. In particular, although the potential of GUI design based on personal characteristics has been studied by customizing the display to meet specific demands [6], there is little empirical data to provide solid guidelines for practitioners to leverage personality variables in this domain.

Table I pinpoints the state-of-the-art research regarding the targeted personality traits, graphical features, and metrics used to study the effect of personality-based GUIs. As we mentioned at the beginning of this section, researchers only addressed an interface elements subset, and not all elements have defined design guidelines. Of the contributors, only Arockiam and Selvaraj [24] provide design guidelines for extravert and neurotic learners in terms of Font Family and Theme (Text Color). As such, there is a considerable lack of research. Although there are several quality dimensions like perceived usability and performance, user preference has been the most studied dimension. In particular, Karsvall [25] and Abrahamian et al. [26] found that participants preferred an interface designed for their personality type. Nevertheless, both studies fall on the mentioned pitfall of designing interfaces beforehand without any input from a sample of participants. As we have already discussed, this may lead to biased results due to the choice of the researchers.

Most studies that found measurable effects with personality traits focus on extraversion. Moreover, a smaller percentage leveraged neuroticism, conscientiousness, and dichotomies from the Myers-Briggs Type Indicator (MBTI). Although Sarsam and Al-Samarraie [7] use the openness to experience and agreeableness traits, these factors were not relevant to differentiate the subjects. As such, there are no studies regarding the effect of these last two traits on user preferences in the context of GUIs. Additionally, the majority of the studies (66.67%) in Table I focuses on personality profiles composed of a unique trait. In contrast, Abrahamian et al. [26], Su et al. [28], and Sarsam and Al-Samarraie [7] leverage more than one personality factor to differentiate users. All the mentioned research gaps allow us to conclude that the state-of-the-art presents several open challenges. In this work, we want to contribute to the state-of-the-art and, at the same time, study the different personality profile compositions for user preference assessment. The following two sections present our work regarding this topic, including personality profiles with a unique factor and multiple factors, as well as how designing GUIs based on user preferences from different personality profiles influences how users perceive those interfaces.

III. Data Collection

As we mentioned in the previous section, our work focuses on studying and incorporating personality traits in GUI design, in particular to extend the current methods of design research and commercial communities, contributing to the HCI research field. Thus, we formulate our research question as: *Are personality-based user preferences a relevant factor to the design of GUIs?* In order to study this effect, we started by choosing which features and their styles we want to address.

A. Design Elements

The core of graphical features we target in this study is similar to Sarsam and Al-Samarraie [9]. In particular, both our work and Sarsam and Al-Samarraie [9] cover information structure, layout type, font style attributes, text size, buttons, color, information density, support, and alignment. While Sarsam and Al-Samarraie [9] asked participants to assess each component of the HSB color model and a set of hues individually, we provide a set of color palettes that we created based on the work of Condeço [21]. We believe that providing the user with optional full color palettes provides a clearer, more rational choice since the user has full information about the final set of colors. This design decision is in contrast to Sarsam and Al-Samarraie [9], where the authors derived the interface color theme from the preference rates of each hue and how much saturated and bright subjects like to see colors. The major limitation of this approach is that the subject cannot see beforehand how the final color palette will look like. This means that there may be some interaction effects between colors in the derived design guidelines that were overlooked by the participants and may have an effect on their interaction.

Sarsam and Al-Samarraie [9] also addressed navigation and list elements. Nevertheless, we do not cover them in this study, since we believe that these graphical elements are better suited for a mobile setting rather than a website desktop-based layout. Regarding navigation, our website desktop-based setting does not have the limited screen size of the typical mobile setting paired with the nonexistence of actions such as hover events. Website desktop-based GUIs have larger screen sizes that allow designers to focus on other design features, as well as support pointer events that would overlap with the design proposals of Sarsam and Al-Samarraie [9]. A similar case can be made for listing elements since its importance is exacerbated by the small screen size that is common in mobile devices. In the webpage desktop-based setting, lists are relegated to an importance similar to other graphical elements, such as images or tables, given that usually the screen size is larger and able to display a larger volume of information. In this case, we decided not to cover this type of design elements as a means to control the complexity of the data analysis.

Besides the nine graphical features that our work shares with Sarsam and Al-Samarraie [9], we decided to also approach three other design elements: body margin, menu structure, and text highlights. Regarding body margin, we believe it is important to assess in the website desktop-based context; it allows us to explore and manipulate more in depth information

TABLE I: Collection of studies focused on user preferences for GUI features influenced by personality traits. The rightmost group includes the quality dimensions used to test the effect of the personality traits.

	Personality Traits							**Graphical Features**											**Quality Metrics**			
	Neuroticism	Extraversion	Openness to Experience	Agreeableness	Conscientiousness	Sensing/Intuition	Thinking/Feeling	Buttons	Element Style	Font Family	Icons	Information Density	Layout	Menu Structure	Navigation	Text Alignment	Text Size	Theme	Mental workload	Perceived usability	User Preference	Performance
Karsvall [25]		x							x									x			x	
Abrahamian et al. [26]		x				x						x									x	
Arockiam and Selvaraj [24]	x	x								x								x				x
Kim et al. [27]		x												x							x	x
Su et al. [28]						x	x					x			x							x
Condeço [21]		x																x			x	
Sarsam and Al-Samarraie [7]	x	x			x			x			x		x	x		x	x	x			x	
Xavier [15]	x														x	x	x		x	x	x	

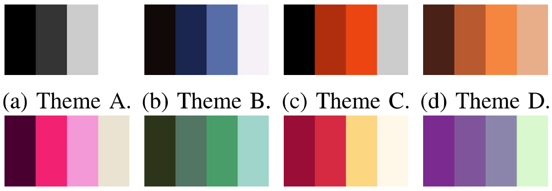
density on the screen, given that the body margin supports the use of white space on the outer border of the main content to change the volume of information on the screen. Menu structure is also an important feature, since it has already shown significant interaction effects with a personality trait in Kim et al. [27]. Since Kim et al. [27] only focused on extraversion, we believe that we can extend their work by including all personality traits from the FFM. Finally, text highlights are common in website desktop-based settings to showcase important information. Alas, we found no study in our research field that leverages this design element. Again, we believe that we can extend the state-of-the-art by including this graphical feature in our study.

Overall, we address low-level text properties (font size, font family, highlights, and text alignment), webpage-level content organization (layout, information density, body margins, and theme), webpage-level organization (menu and information structures) and, finally, other elements such as buttons and support. We believe that these elements include the GUI features more frequently mentioned in the literature as well as the most relevant levels of context granularity in webpage-based GUIs. Although some of these features are less abstract, such as font size or family, elements such as information density and structure have different styles that are harder to visually exemplify. In order to bridge this gap, we include with this kind of features a brief explanation regarding what is their meaning and how they are present in a website layout. The elements are described as follows:

- **Body margins**: It addresses the space between the main content and the limits of the GUI. We tested for small, medium, and large margins.
- **Buttons**: A graphical element that can be clicked to prompt an action. Our study focuses on three types of buttons: buttons with a name, buttons with an icon, and buttons with a name and an icon.
- **Information density**: It denotes the volume of graphical and textual elements in the display. We presented three different amounts of information density: low, medium, and high information density.
- **Information structure**: It refers to the organization of data in the GUI. We decided to focus on four different settings: linear structure, hierarchical structure, network structure, and matrix structure [9].
- **Layout type**: It refers to the arrangement of the GUI components. Similar to Sarsam and Al-Samarraie [9], we focus on linear, relative, and web view layouts.
- **Menu structure**: We focus on the depth and breadth dimensions in the menu structure design.
- **Support**: It provides hints to the user that are embedded usually within the design of GUIs. Likewise Sarsam and Al-Samarraie [9], we test support items based on icons and text.
- **Text alignment**: It refers to how information is arranged within compartments. We study justified, left, and center alignments.
- **Text font**: It refers to the font family of the text. We study several font families, namely Arial, Courier, Georgia, Handwritten, Times New Roman, and Verdana.
- **Text highlights**: It establishes how relevant information is highlighted. We use background color, bold, and underline styles.
- **Text size**: It refers to the size of the text compared to the size of the GUI. We tested for small, medium, and large font sizes.
- **Theme**: The theme of the GUI has a set of colors to use on the drawing of the elements. Based on Condeço [21], the selection of colors was in accordance to hue, saturation, and brightness (Figure 1).

B. Personality Data

Regarding personality variables, we use the personality traits and their facets from the Five-Factor Model (FFM). The FFM is the most widespread and generally accepted model of personality [8], [29], [30], since it provides a nomenclature and a conceptual framework that unifies much of the research



(a) Theme A. (b) Theme B. (c) Theme C. (d) Theme D.

(e) Theme E. (f) Theme F. (g) Theme G. (h) Theme H.

Fig. 1: The different styles for the theme feature.

findings in the psychology of individual differences[1]. This model consists of five general traits to describe personality and 30 facets of personality as follows:

- **Neuroticism** (*Anxiety* (N1), *Anger* (N2), *Depression* (N3), *Self-consciousness* (N4), *Immoderation* (N5), *Vulnerability* (N6)): distinguishes the stability of emotions and even-temperedness from negative emotionality, which can be described as feeling nervous, sad, and tense [32]. It is often referred to as emotional instability, addressing the tendency to experience mood swings and negative emotions such as anxiety, worry, fear, anger, frustration, envy, jealousy, guilt, depressed mood, and loneliness [33].
- **Extraversion** (*Friendliness* (E1), *Gregariousness* (E2), *Assertiveness* (E3), *Activity level* (E4), *Excitement-seeking* (E5), *Cheerfulness* (E6)): suggests a lively approach toward the social and material world [32]. It measures a person's tendency to seek stimulation in the external world, the company of others, and to express positive emotions.
- **Openness to experience** (*Imagination* (O1), *Artistic interests* (O2), *Emotionality* (O3), *Adventurousness* (O4), *Intellect* (O5), *Liberalism* (O6)): describes the wholeness and complexity of an individual's psychological and experiential life [32]. It measures a person's imagination, curiosity, seeking of new experiences, and interest in culture, ideas, and aesthetics. It is related to emotional sensitivity, tolerance, and political liberalism.
- **Agreeableness** (*Trust* (A1), *Morality* (A2), *Altruism* (A3), *Cooperation* (A4), *Modesty* (A5), *Sympathy* (A6)): distinguishes pro-social and communal orientation toward others from antagonism [32]. It measures the extent to which a person is focused on maintaining positive social relations.
- **Conscientiousness** (*Self-efficacy* (C1), *Orderliness* (C2), *Dutifulness* (C3), *Achievement-striving* (C4), *Self-discipline* (C5), *Cautiousness* (C6)): suggests self-use of socially prescribed restraints that facilitate goal completion, following norms and rules, and prioritizing tasks [32]. It measures the preference for an organized approach to life as opposed to a spontaneous one.

[1]Several personality researchers agree that these five personality traits are representative of cross-cultural individual differences in normal behavior and studies have replicated this taxonomy in a diversity of samples [31].

C. Apparatus

The native version of the Revised NEO Personality Inventory (NEO PI-R) [34] was developed by Lima and Simões [35] to assess personality variables from the FFM. The NEO PI-R has a high internal consistency with values ranging from 0.79 to 0.86 [35]. It has 240 items and allows researchers to assess the FFM five personality traits and their 30 facets. The questionnaire identifies the intensity of each personality trait of a person using high-score and low-score features. The questionnaire has 30 different subscales (one for each facet), with eight items for each subscale. Thus, every trait has 48 different items. Additional experimental setup included an online questionnaire with the features and their different styles to assess user preferences[2]. Each style was accompanied by an illustrative image and an explanation.

D. Procedure

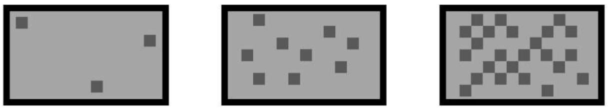
We recruited subjects through standard convenience sampling procedures by direct contact and word of mouth. Subjects included any native interested in participating with at least 18 years old. Our data set comprises 65 participants (31 males, 34 females) between 18 and 60 years old ($M = 24.03; SD = 6.81$). All participants had a normal or corrected-to-normal vision, and there were no color blind subjects as assessed by a validated simplified version of the Ishihara test [36]. Additionally, we found that the apparatus (mobile, desktop, or tablet) through which participants assessed their design preferences did not lead to statistically significant differences in their ratings.

Before the experiment, participants were informed about the experience and invited to agree with a compulsory consent form. We also informed them that they could quit the experiment at any time. Beforehand, participants filled in the NEO PI-R in an online platform to collect the personality traits and their facets from the FFM. Afterward, we invited participants to fill in the online questionnaire that presented all features and visual examples of the styles in a fixed order to assess user preference for each arrangement. In particular, each participant assessed their preference for a style of a feature by completing a seven-point Likert scale ranging from *Low Preference* (1) to *High Preference* (7). As an example, Figure 2 presents the set of possible styles for the information density feature (see the supplemental material for the full questionnaire). Finally, participants received their compensation.

IV. Data Analysis

This section describes how we created the design guidelines for the different personality profiles. It starts by analysing how to cluster personality characteristics, and continues by exploring association rules from patterns regarding user preferences. In particular, we conducted a mixed analysis, where the within-subjects variables are the ratings that each participant

[2]https://drive.google.com/file/d/14jgCLIRCcixXT_Dn1gTRaC_EEXuCCZaL/view?usp=sharing (Last access: September 30, 2022).



(a) Low density. (b) Medium density. (c) High density.

Fig. 2: The different styles for the information density feature (adapted from Sarsam and Al-Samarraie [7]).

attributed to the styles and the between-subjects variables are the personality traits and their facets from the FFM.

A. *Clustering Personality Characteristics*

There are several ways to understand how personality models design choice preferences. One approach is to first convert each personality variable to categorical values following either the quartile distributions of the sample or the native population [35], and then analyse each personality variable separately using ANOVAs to explore main and interaction effects with the features and its styles according to user preferences. The other approach is by clustering users according to their personality characteristics and find whether participants with similar personality profiles share preferences for certain GUI elements. In our work, we focus on the second approach. Although it is also possible to categorize personality variables as aforementioned and then cluster users, we decided to work with integers, as they allow a finer granularity compared to data binning. In particular, we use the 30 facets from the FFM as input variables for the clustering algorithms.

We started by applying hierarchical clustering [37] to find the most appropriate number of clusters to work with. In particular, we allow the algorithm to choose the minimal cluster size (the smallest size grouping that we wish to consider a cluster) and how conservative the algorithm should be while clustering (the number of points that are declared as noise) according to the best silhouette and Davies–Bouldin index scores [38]. Additionally, we use the euclidean distance as the clustering metric since we are working with integer values. This approach allows the algorithm to search in a given set of parameter values which combination of arguments generates the best clustering solution according to the silhouette and Davies–Bouldin index scores. Following this approach, hierarchical density-based clustering [39] yielded three clusters. We followed up with the k-means clustering algorithm [40] as a way to avoid the noise labels that hierarchical density-based clustering produces. By fixing the number of clusters to three, we normalized the data and allowed k-means to run 100 times with different centroid seeds using Euclidean distance. The final result contained the best output of 100 consecutive runs in terms of inertia.

The distributions of personality variables from the clusters are presented in Figure 3. The first cluster ($N = 20$) notably has participants with the highest levels of extraversion ($M = 124.10; SD = 14.46$) and openness to experience ($M = 136.20; SD = 16.17$), followed by medium levels of neuroticism ($M = 96.15; SD = 17.36$) and conscientiousness ($M = 119.05; SD = 16.04$), and low levels of agreeableness ($M = 119.05; SD = 18.87$). For simplicity, we labeled this cluster as "Extraversion-Openness" (C-EO), as those traits present the highest means compared with other clusters. In contrast, the second cluster ($N = 19$) shows high neuroticism ($M = 121.47; SD = 16.76$), and low extraversion ($M = 95.74; SD = 17.46$), openness to experience ($M = 118.79; SD = 16.65$), agreeableness ($M = 125.42; SD = 12.99$), and conscientiousness ($M = 108.84; SD = 19.89$). Therefore, we labeled this cluster as "Neuroticism" (C-N). Finally, the third cluster ($N = 26$) includes participants with high agreeableness ($M = 132.38; SD = 16.29$) and conscientiousness ($M = 135.54; SD = 16.18$), medium levels of extraversion ($M = 107.38; SD = 16.99$), and low levels of neuroticism ($M = 80.12; SD = 12.94$) and openness to experience ($M = 115.04; SD = 16.30$). We labeled it as "Agreeableness-Conscientiousness" (C-AC) since this cluster shows higher scores for agreeableness and conscientiousness. Although most traits follow the native distribution [35], the dimensions of openness to experience and conscientiousness show higher and lower medians, respectively. These differences may be due to the sampling of our study being composed of young adults (from 18 to 24 years old) and adults (older than 24) from a university setting. Nevertheless, the interquartile range (IQR) shows a well-balanced distribution for these cases.

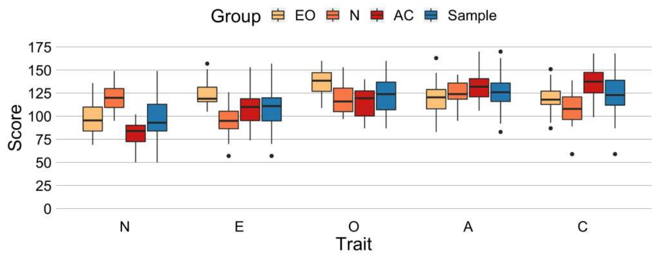


Fig. 3: Boxplots of the distribution of traits between clusters and the sample.

Additionally, we conducted an ANOVA to validate whether the clusters are interdependent regarding personality traits. We found a significant difference in neuroticism ($F(2, 62) = 38.938, p < .001$), extraversion ($F(2, 62) = 14.812, p < .001$), openness to experience ($F(2, 62) = 10.198, p < .001$), agreeableness ($F(2, 62) = 3.823, p = .027$), and conscientiousness ($F(2, 62) = 13.7, p < .001$) across the three clusters. There were also significant differences in 17 personality facets out of 30. These results show that clusters differ in many personality scores, notably at a trait level. Therefore, each of the three clusters contains a different stable and valid user group than the other clusters. In particular, C-EO contains people that are outgoing, talkative, and show an energetic behavior open to new experiences. In contrast, C-N depicts people that are not emotionally stable and have a tendency to experience mood swings. Finally, C-AC includes pro-social people that focus on maintaining positive social relations while following socially prescribed restraints to have an organized approach to life. After identifying the different personality groups, our next objective is to extract design guidelines among individuals of those three clusters. We used an association rules method to identify the design preferences for each personality profile.

B. Extracting Association Rules

We used the Apriori algorithm [41] to find common patterns between the preferred styles of each participant. We started by creating an array containing the style preferred the most for each feature per user. In case of ties in the preference rate between styles, we included all arrangements tied together. For instance, if the subject rated their preference for "medium" and "high" information density with 6 and the "low" with 4, we included both the "medium" and "high" styles in the array. Next, we divided users by their cluster labels and used the Apriori in each cluster. Each run was performed with lower bound minimal values of 0.15 for support, 0.9 for confidence, and 6 for lift. We empirically tested these values to reach a core of rules as robust as possible. The algorithm yielded 74 rules for the C-EO, 282 for the C-N, and 6 for the C-AC.

C. Finding Preferences for Clusters

We continued our analysis by choosing which rules to use on the design guidelines according to the frequency of each rule. We started by choosing the rule with the highest frequency value and then continued by picking rules with lower frequency that share a design style and do not conflict with a design style previously selected for a feature. In addition, we focused on maximizing the number of design elements that could be derived from the association rules. When a feature did not have a style associated with it at the end of our analysis, we chose the most frequent preferred style for that feature among participants of the cluster. Table II illustrates the final rule sets for each cluster. Based on the final set of rules for each cluster, we were able to derive which styles to apply to the different GUI elements (Table III). There are several features that have different styles across versions: font family and size, information density, layout, text align, and theme. Nevertheless, we were not able to derive styles for certain features.

As we mentioned, we address this issue by choosing the most frequent style among cluster participants. Regarding the C-EO, only the styles of *Highlights* and *Information Structure* features were not derived from the association rules. The most common styles were "bold" and "hierarchy", respectively. Moreover, both the "bold" highlights (M = 5.90; SD = 1.07) and the "hierarchy" information structure (M = 5.85; SD = 0.88) were favored by the participants. For the C-N, *Information Structure* was the only feature assessed by the post-analysis based on the frequency of styles, resulting in the "hierarchy" style that also yielded positive ratings in design preference (M = 5.95; SD = 0.78). Finally, we derived five features' styles from them though the C-AC had only three defined rules. The remaining features were *Font Size*, *Help*, *Highlights*, *Information Density*, *Information Structure*, *Menu*, and *Theme*, which yielded the positively rated styles of "medium" (M = 5.65; SD = 1.26), "icon" (M = 5.46; SD = 1.07), "bold" (M = 5.85; SD = 0.92), "medium" (M = 4.96; SD = 1.31), "hierarchy" (M = 5.73; SD = 0.96), "breadth" (M = 5.81; SD = 0.90), and the mono-chromatic blue theme (M = 5.31; SD = 1.32), respectively.

TABLE II: Association rules chosen for each cluster. An association rule from the Apriori algorithm is often represented as $styleA \rightarrow styleB$, which translates into $styleB$ being frequently present in a set of preferences that also contains $styleA$.

Rules for the C-EO	**Frequency**	**Support**	**Confidence**	**Lift**
themeB → layoutRelative	13	0.150	1.00	6.667
layoutRelative → menuBreadth	7	0.150	1.00	6.667
buttonIconText → menuBreadth	6	0.150	1.00	6.667
themeB → menuBreadth	5	0.150	1.00	6.667
marginSmall → menuBreadth	4	0.150	1.00	6.667
marginSmall → alignJustified	3	0.150	1.00	6.667
buttonIconText → alignJustified	2	0.150	1.00	6.667
densityMedium → layoutRelative	2	0.150	1.00	6.667
buttonIconText → densityMedium	2	0.150	1.00	6.667
themeB → buttonIconText	2	0.150	1.00	6.667
densityMedium → buttonIconText	1	0.150	1.00	6.667
layoutRelative → sizeLarge	1	0.150	1.00	6.667
menuBreadth → sizeLarge	1	0.150	1.00	6.667
themeB → sizeLarge	1	0.150	1.00	6.667
marginSmall → helpIcon	1	0.150	1.00	6.667
buttonIconText → marginSmall	1	0.150	1.00	6.667
Rules for the C-N	**Frequency**	**Support**	**Confidence**	**Lift**
themeA → marginSmall	29	0.158	1.00	6.333
themeA → buttonIconText	15	0.158	1.00	6.333
densityLow → themeA	9	0.158	1.00	6.333
layoutLinear → themeA	4	0.158	1.00	6.333
themeA → menuBreadth	3	0.158	1.00	6.333
marginSmall → densityLow	2	0.158	1.00	6.333
marginSmall → layoutLinear	1	0.158	1.00	6.333
layoutLinear → highlightBold	1	0.158	1.00	6.333
themeA → helpIcon	1	0.158	1.00	6.333
Rules for the C-AC	**Frequency**	**Support**	**Confidence**	**Lift**
alignLeft → marginSmall	4	0.160	1.00	6.250
alignLeft → buttonIconText	1	0.160	1.00	6.250
layoutRelative → fontGeorgia	1	0.160	1.00	6.250

With the features and their styles defined for each cluster, we can create personality-based GUI design guidelines for different elements. In particular, we were able to derive the following guidelines:

- *People high on extraversion and openness to experience prefer GUIs with large Arial font, medium information density, relative layout, justified text, and a mono-chromatic blue theme.*
- *People high on neuroticism prefer GUIs with medium Arial font, low information density, linear layout, justified text, and a gray-scale theme.*
- *People high on agreeableness and conscientiousness prefer GUIs with medium Georgia font, medium information density, relative layout, left-align text, and a mono-chromatic blue theme.*

As our findings show similar preferences for button, help, highlights, menu, and structure types, in addition to the size of the margins in a GUI, preferences regarding these design features may be independent of personality traits. Indeed,

TABLE III: Features and preferred styles for each cluster. The percentage represents the amount of times the design style was chosen compared to the other styles for a feature in each cluster. Bold styles were derived from the association rules. Highlighted rows present differences in styles among distinct personality groups.

Feature	C-EO	C-N	C-AC
Buttons	**IconText** (48%)	**IconText** (43%)	**IconText** (48%)
Font Family	**Arial** (34%)	**Arial** (45%)	**Georgia** (13%)
Font Size	**Large** (38%)	**Medium** (50%)	Medium (69%)
Help	**Icon** (71%)	**Icon** (75%)	Icon (63%)
Highlights	Bold (59%)	**Bold** (61%)	Bold (57%)
Information Density	**Medium** (57%)	**Low** (23%)	Medium (55%)
Information Structure	Hierarchy (41%)	Hierarchy (71%)	Hierarchy (43%)
Layout	**Relative** (42%)	**Linear** (48%)	**Relative** (48%)
Margin	**Small** (63%)	**Small** (50%)	**Small** (65%)
Menu	**Breadth** (86%)	**Breadth** (76%)	Breadth (81%)
Text Align	**Justified** (87%)	**Justified** (85%)	**Left** (24%)
Theme	**B** (34%)	**A** (20%)	B (29%)

a closer look at Table III shows that the three personality profiles often preferred the same style for the GUI features, thus suggesting that the groups may not have differed much regarding those elements.

V. Discussion and Future Work

Our findings add to prior work by Alves et al. [3], who found that there is a strong need for personality-based GUI design research, and to the work of Sarsam and Al-Samarraie [9], who addressed a subtopic of GUI design by focusing on mobile applications. Similar to Arockiam and Selvaraj [24], we found that extraversion and neuroticism have an effect on the font family, since users with high values on both traits prefer Arial font, while people with lower values tend to prefer Georgia. We also found that conscientiousness, extraversion, and neuroticism have an effect on how people prefer the size of the text on the screen [7], [15], since people with high extraversion prefer large fonts, while the remaining would rather have medium font size. Information density also showed differences between personality profiles [26], with people with high neuroticism preferring lower densities. Additionally, we found that people with high neuroticism prefer themes in a gray scale, and people with on the other traits would rather have a monochromatic blue theme varying in value, according to the HSV color scheme. In particular, we found contradictory results compared to the work of Condeço [21] by showing that introverts prefer gray-scale themes and extraverts blue tones. We also found contrary results compared to Sarsam and Al-Samarraie [7] and Kim et al. [27], since there were no differences regarding buttons and menu structure. The next step in our research is to validate the design guidelines. In particular, we want to design GUIs according to the preferences of each group and then examine how the different personality profiles interact with those interfaces.

Some relevant factors may explain the lack of significance observed in some results. First, the number of participants in this experiment could have been higher, as a higher number of participants would allow conclusions with a better impact. Our results are limited to native users and may not transfer to other populations. Both factors are relevant when considering a personality profile with a level of complexity based on five personality traits. Although this variation may lead to different preferences from each cluster, our methodology is sound to differentiate people based on personality factors since each group was interdependent from others for all personality traits. Another common point to the previous study is that we could show more styles to participants that may have revealed other preferences. However, we chose the most common styles for the features. In addition, basing our approach on association rules and shared patterns of design preferences may hinder design styles that are most frequently chosen as the preferred one independently of their relationship with the remaining features. In other words, although an association rule may be relatively more frequent, it does not mean that it applies to the cluster as a whole. Thirdly, the images used to illustrate the different styles may have affected how people perceived them. Given its abstract nature, people may have over-fitted their preference regarding certain element styles and, therefore, assess their choice based on one particular experience instead of assuming a general scenario. Further, the design guidelines assume that the personality of users must be known beforehand. Although some studies have already been able to predict personality characteristics without questionnaires [42], further research is needed. Finally, the design guidelines were not validated. Future studies should include a validation of the guidelines in specific design contexts to understand how they can be applied to develop better user interfaces in practice.

VI. Conclusion

Our objective was to assess how personality variables model user preferences. We focused on the FFM to represent people and used the NEO PI-R to model their personality variables. Moreover, we addressed the most used GUI elements of the state-of-the-art and based their style variations on past research. Our approaches aggregated users based on their personality variables to then extract design preferences from various personality profiles. On the basis of an association rules technique, results showed that different personality profiles have distinct preferences for certain GUI element styles. Notably, although our objective is not to validate the personality profiling, we were able to identify the design preferences of three different personality profiles that are well-suited to separate the population according to the five personality traits of the FFM. Additionally, we identified which features are independently modelled by user preferences.

Acknowledgment

This work was supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with references SFRH/BD/144798/2019 and UIDB/50021/2020.

References

[1] Z. Liu, R. Crouser, and A. Ottley, "Survey on individual differences in visualization," *Computer Graphics Forum*, vol. 39, pp. 693–712, 06 2020.

[2] Y. Dwivedi, N. Rana, A. Jeyaraj, M. Clement, and M. Williams, "Re-examining the unified theory of acceptance and use of technology (utaut): Towards a revised theoretical model," *Information Systems Frontiers*, vol. 21, pp. 1–16, 06 2019.

[3] T. Alves, J. Natálio, J. Henriques-Calado, and S. Gama, "Incorporating personality in user interface design: A review," *Personality and Individual Differences*, vol. 155, p. 109709, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S019188691930649X

[4] O. Kernberg, "What is personality?" *Journal of personality disorders*, vol. 30, pp. 145–156, 05 2016.

[5] A. M. Viveros, E. Hernández Rubio, and D. E. Vázquez Ceballos, "Equivalence of navigation widgets for mobile platforms," in *Design, User Experience, and Usability. User Experience Design for Diverse Interaction Platforms and Environments*, A. Marcus, Ed. Cham: Springer International Publishing, 2014, pp. 269–278.

[6] Y. Karanam, L. Filko, L. Kaser, H. Alotaibi, E. Makhsoom, and S. Voida, "Motivational affordances and personality types in personal informatics," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, ser. UbiComp '14 Adjunct. New York, NY, USA: ACM, 2014, pp. 79–82. [Online]. Available: http://doi.acm.org/10.1145/2638728.2638800

[7] S. M. Sarsam and H. Al-Samarraie, "Towards incorporating personality into the design of an interface: a method for facilitating users' interaction with the display," *User Modeling and User-Adapted Interaction*, vol. 28, no. 1, pp. 75–96, Mar 2018. [Online]. Available: https://doi.org/10.1007/s11257-018-9201-1

[8] P. Costa and R. R. McCrae, "The revised neo personality inventory (neo-pi-r)," *The SAGE Handbook of Personality Theory and Assessment*, vol. 2, pp. 179–198, 01 2008.

[9] S. M. Sarsam and H. Al-Samarraie, "A first look at the effectiveness of personality dimensions in promoting users' satisfaction with the system," *SAGE Open*, vol. 8, no. 2, p. 2158244018769125, 2018. [Online]. Available: https://doi.org/10.1177/2158244018769125

[10] N. Makris and M. van Eekelen, "Creating adaptable and adaptive user interface implementations in model driven developed software," MSc dissertation, Radboud University, Nijmegen, Netherlands, 2016.

[11] J. Kim, "Modeling task-based information seeking on the web: Application of information seeking strategy schema," *Proceedings of the American Society for Information Science and Technology*, vol. 44, pp. 1 – 13, 10 2008.

[12] H. Al-Samarraie, A. Eldenfria, and H. Dawoud, "The impact of personality traits on users' information-seeking behavior," *Information Processing & Management*, vol. 18, pp. 17–18, 10 2016.

[13] V. Kostov and S. Fukuda, "Development of man-machine interfaces based on user preferences," in *Proceedings of the 2001 IEEE International Conference on Control Applications (CCA'01) (Cat. No.01CH37204)*, Sep. 2001, pp. 1124–1128.

[14] M. Chae and J. Kim, "Do size and structure matter to mobile users? an empirical study of the effects of screen size, information structure, and task complexity on user activities with standard web phones," *Behaviour & Information Technology*, vol. 23, no. 3, pp. 165–181, 2004. [Online]. Available: https://doi.org/10.1080/01449290410001669923

[15] I. Xavier, "Neural: Towards neuroticicism-based user interface customization," MSc dissertation, Universidade de Lisboa - Instituto Superior Técnico, 2019.

[16] S. Basu, *Modern UI Design*. Apress, 05 2013, pp. 11–24.

[17] L. Evett and D. Brown, "Text formats and web design for visually impaired and dyslexic readers—clear text for all," *Interacting with Computers*, vol. 17, no. 4, pp. 453 – 472, 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0953543805000378

[18] D.-S. Lee, K.-K. Shieh, S.-C. Jeng, and I.-H. Shen, "Effect of character size and lighting on legibility of electronic papers," *Displays*, vol. 29, no. 1, pp. 10 – 17, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0141938207000492

[19] S. K. Kane, J. O. Wobbrock, and I. E. Smith, "Getting off the treadmill: Evaluating walking user interfaces for mobile devices in public spaces," in *Proceedings of the 10th International Conference on Human Computer Interaction with Mobile Devices and Services*, ser. MobileHCI '08. New York, NY, USA: ACM, 2008, pp. 109–118. [Online]. Available: http://doi.acm.org/10.1145/1409240.1409253

[20] K. Reinecke and A. Bernstein, "Knowing what a user likes: A design science approach to interfaces that automatically adapt to culture," *MIS Quarterly*, vol. 37, pp. 427–453, 02 2013.

[21] J. Condeço, "Colorcode: Exploring social and psychological dimensions of color," MSc dissertation, Universidade de Lisboa - Instituto Superior Técnico, 2018.

[22] J. Ribeiro and M. Carvalhais, "Web design patterns for mobile devices," in *Proceedings of the 19th Conference on Pattern Languages of Programs*, ser. PLoP '12. USA: The Hillside Group, 2012, pp. 13:1–13:48. [Online]. Available: http://dl.acm.org/citation.cfm?id=2821679.2831283

[23] J. GeiBler, M. Gauler, and N. A. Streitz, "Evaluating gedrics: usability of a pen-centric interface," in *Human-computer Interaction, INTERACT'99: IFIP TC. 13 International Conference on Human-Computer Interaction, 30th August-3rd September 1999, Edinburgh, UK*, vol. 1. IOS Press, 1999, p. 222.

[24] L. Arockiam and J. C. Selvaraj, "User interface design for effective e-learning based on personality traits," *International Journal of Computer Applications*, vol. 61, pp. 28–32, 01 2013.

[25] A. Karsvall, "Personality preferences in graphical interface design," in *Proceedings of the Second Nordic Conference on Human-Computer Interaction*, ser. NordiCHI '02. New York, NY, USA: Association for Computing Machinery, 2002, p. 217–218. [Online]. Available: https://doi.org/10.1145/572020.572049

[26] E. Abrahamian, J. Weinberg, M. Grady, and C. Michael Stanton, "The effect of personality-aware computer-human interfaces on learning." *J. UCS*, vol. 10, pp. 17–27, 01 2004.

[27] J. Kim, A. Lee, and H. Ryu, "Personality and its effects on learning performance: Design guidelines for an adaptive e-learning system based on a user model," *International Journal of Industrial Ergonomics*, vol. 43, p. 450–461, 09 2013.

[28] K.-W. Su, C.-J. Chen, and L.-Y. Shue, "Implication of cognitive style in designing computer-based procedure interface," *Human Factors and Ergonomics in Manufacturing & Service Industries*, vol. 23, no. 3, pp. 230–242, 2013. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/hfm.20315

[29] L. Goldberg, "The structure of phenotypic personality traits," *American Psychologist*, vol. 48, pp. 26–34, 02 1993.

[30] M. T. Russell, R. B. Cattell, A. Cattell, H. E. Cattell, and D. L. Karol, *16PF fifth edition administrator's manual*. Institute for Personality and Ability Testing, Incorporated, 1994.

[31] T. Chamorro-Premuzic and A. Furnham, *Personality and intellectual competence*. Psychology Press, 2014.

[32] S. Halko and J. A. Kientz, "Personality and persuasive technology: an exploratory study on health-promoting mobile applications," in *International conference on persuasive technology*. Springer, 06 2010, pp. 150–161.

[33] E. R. Thompson, "Development and validation of an international english big-five mini-markers," *Personality and Individual Differences*, vol. 45, no. 6, pp. 542 – 548, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0191886908002195

[34] P. T. Costa Jr and R. R. McCrae, *The Revised NEO Personality Inventory (NEO-PI-R)*. Sage Publications, Inc, 2008.

[35] M. Lima and A. Simões, "Neo-pi-r manual profissional," *Lisboa: CEGOC*, 2000.

[36] D. V. de Alwis and C. H. Kon, "A new way to use the ishihara test," *Journal of neurology*, vol. 239, no. 8, pp. 451–454, 1992.

[37] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*. Elsevier, 2011.

[38] S. Petrovic, "A comparison between the silhouette index and the davies-bouldin index in labelling ids clusters," in *Proceedings of the 11th Nordic Workshop of Secure IT Systems*, 2006, pp. 53–64.

[39] L. McInnes, J. Healy, and S. Astels, "hdbscan: Hierarchical density based clustering," *Journal of Open Source Software*, vol. 2, no. 11, p. 205, 2017.

[40] G. Zeng, "Fast approximate k-means via cluster closures," in *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, ser. CVPR '12. USA: IEEE Computer Society, 2012, p. 3037–3044.

[41] M. Ilayaraja and T. Meyyappan, "Mining medical data to identify frequent diseases using apriori algorithm," in *2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering*. IEEE, 2013, pp. 194–199.

[42] N. Majumder, S. Poria, A. Gelbukh, and E. Cambria, "Deep learning-based document modeling for personality detection from text," *IEEE Intelligent Systems*, vol. 32, no. 2, pp. 74–79, 2017.

Exploring how Temporal Framing Affects Trust with Time-series Visualizations

Tomás Alves, Carlota Dias, Daniel Gonçalves, Sandra Gama
HUMAN Lab, Graphics and Interaction, INESC-ID, IST,
University of Lisbon, Portugal
{tomas.alves,carlota.lopes.dias,daniel.j.goncalves,sandra.gama}@tecnico.ulisboa.pt

***Abstract*—Trust is one of the most relevant factors when users build knowledge from visualization to predict whether they will use the represented information. In particular, trust perception is the user's subjective evaluation of the quality and reliability of the visualized information. However, research leveraging information visualization techniques to study trust perception is limited. This work studies whether varying the temporal framing of line charts affects trust perception in an uncertain scenario. Our results suggest that granularity may be relevant for time-based visualization design. In particular, individuals trust more in a line chart with a higher number of data points and interact more with a line chart in which they trust less. These findings contribute to the state-of-the-art research in visual analytic systems by empowering designers to understand how trust perception in a health emergency scenario varies for line charts with different temporal frames.**

***Index Terms*—information visualization, trust, user interaction, time-series**

I. INTRODUCTION

The core purpose of an Information Visualization (InfoVis) is to help users discover, explain, and form decisions based on the information that is conveyed [1]. Visualizations may follow several strategies to aid decision-making, e.g., offering sufficient guidance [2] or emphasizing critical information [3]. However, InfoVis research shows that visualization-supported decision-making vulnerable to cognitive biases and uncertainty [4], [5] Additionally, other human factors affect how the user derives a choice of direction as an outcome. For instance, when users interact with a visualization, the level of trust that users have in a new knowledge affects their decision-making process [6].

In InfoVis, trust defines the tendency of a user to rely on visualization and build on the information displayed [7]. In particular, trust becomes particularly important when there is some risk associated with the information; it allows the user to minimize the uncertainty belonging to digital data, especially when they are vulnerable to suffering a loss if they believe in the information displayed [7]. However, trust remains a challenge in InfoVis design [8] since there is still limited evidence regarding what might lead a user to trust in visualization without an extensive elaboration of the information [7].

In our study, we focus on trust perception in an uncertainty scenario. Further, we can assume that when users are prompt to decide with risk, they will elaborate on the trustworthiness of the visualization more in-depth yet rely on superficial trust cues in less relevant or less risky tasks [7]. Inspired by these findings, we place subjects in a decision-making situation with an involved risk. At the same time, this situation is familiar and relatable enough to emphasize the risky nature of the decision. In particular, we use a health emergency scenario when there is an overcrowding crisis, i.e., there is no space left to meet the timely needs of the following patient requiring emergency care [9]. We opted for this topic because we believe that most people are familiar with having to wait for medical support in the emergency department. Then, we ask users to make decisions and perform tasks based on time-oriented information visualizations with varying domain framing factors. The study aimed to address questions such as: Does the temporal frame of the time series affect trust perception? Does the degree of confidence in a choice vary across several decisions? Do users interact more with a visualization they trust the most?

II. RELATED WORK

The topic of trust has been relatively underexplored in visual analytics. Similar to human relations where there are both a trustor and a trustee, trust in the InfoVis context encompasses trustworthiness and trust perception [7]. The trustworthiness of the visualization depends on the characteristics of the visualization like data accuracy, objectivity, and completeness [10]. For instance, Xiong et al. [10] explored whether there is a relationship between trust and data visualization transparency – the perceived quality and quantity of intentionally shared information [11]. Xiong et al. [10] asked participants to put themselves in the role of a firefighter to promote a frame of reference. Then, each participant chose which visualization to use and from which fire station they should dispatch the firefighters. Participants were shown two different visualizations and then told these visualizations were screenshots of several driving applications that displayed the routes from several fire stations to a fire location. These visualizations varied in the displayed volume of information by changing the number of possible routes, the number of fire stations recommended, and the number of fastest paths. Results showed that participants were more likely to choose visualizations that appeared to be clear, more thorough, and disclosed a higher amount of information.

978-8-3503-4573-57/22/$31.00 ©2022 IEEE

Other studies have shown that design factors such as usability and user experience or the amount of processed underlying data can affect trustworthiness [12], [13]. For instance, a recent study by Bartram et al. [14] shows that trust plays a significant role in data workers. In particular, these workers would be willing to perform monotonous and repetitive tasks to maintain immediate access, control, and understanding over their actions and sense-making process.

Regarding trust perception, it tackles the evaluation of the quality and reliability of the visualized information [7]. While Kong et al. [13] suggest that the misalignment of graphical elements affects the understanding and, consequently, the credibility of the information depicted, other researchers imply that prior experiences play a relevant role [15], [16]. For instance, Dasgupta et al. [17] studied the level of trust of domain scientists in visual analytics systems as opposed to more common manual analysis methods. The authors were able to find that, despite being unfamiliar with a visual analytic system, the experts had an average level of trust comparable with the same in conventional analysis methods. The core factors for the analytical system are that it should be intuitive, transparent, and allow a seamless switch between hypothesis generation and evidence gathering. Finally, user intentions and perceived risk may also influence trust perception [18], [19]. These studies collectively show that studying trust in visualization offers an opportunity for the state-of-the-art. Our work builds on the mentioned studies for trust assessment, focusing on whether temporal dimensions affect trust perception.

III. Methodology

Our research question focuses on analyzing **whether temporal framing has an effect on trust perception in the context of time-oriented linear charts of healthcare information**. We leverage the self-assessment of trust and their interaction data with the information visualizations.

A. Temporal Framing

There is a wide range of data features to consider when studying trust perception. Time dimensionality [20] may hold promising results given its relevance in recent research [21], [22]. Time-oriented data includes dimensions such as linear $vs.$ cyclic time, time points $vs.$ time intervals, and order time $vs.$ branching time. Consequently, different time features lead to alternative visualization techniques. Among the different time-based aspects of interest, we believe that *temporal framing*, i.e., the temporal scope that is presented in the domain of visualization, such as daily, weekly, or monthly, may hold promising results. This feature can offer several factors to manipulate, e.g., time range or value aggregation. Therefore, temporal framing may significantly affect the level of detail of the information and, consequently, the granularity of the data or the number of data points.

For instance, Oscar et al. [23] studied the consequences of mismatching the granularity of information presented on visualization to user needs. Participants were shown different visualizations and asked to complete tasks that required information that might not be available in the visualization granularity. In particular, in some cases, users were not presented with enough information to answer the questions. Results showed that when users analyzed information mismatched to the need for detail required by the task, they were less likely to complete the assignment correctly. Moreover, participants were often unable to identify that the visualization did not include the information required to complete the task. Consequently, results demonstrated that using an appropriate visualization is a crucial performance factor. Although the mentioned studies recognize that different visual techniques affect the decisions users make when analyzing distinct visualizations, as far as we know, there is no information concerning how using time-series visualization techniques impacts users' perceived trust.

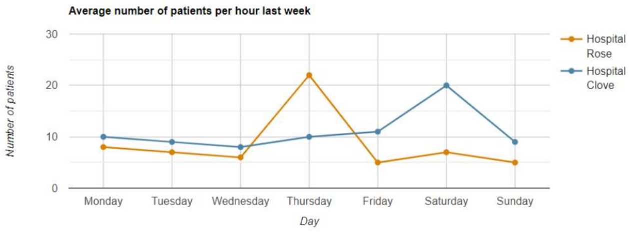
B. Visualizations

Research has shown that positional encodings such as line charts are the best option to visualize time-oriented data on decision-making processes [21], [24]. Moreover, line charts are one of the most well-known visualization idioms, and familiarity with a visualization system inspires trust, whereas novel visualizations may act as a barrier [15]. Based on these findings, we opted to use line charts to depict the variation of a continuous variable along a time axis (Figure 1), as this type of graph is the most common form of time-series visualization [25]. Each line chart displays the number of patients in an emergency room throughout a time in two different hospitals, encoded by two colors randomly associated at the beginning of each experiment. While the x-axis illustrates the time, the y-axis represents the number of patients, and each line presents a hospital. Each line chart has one of the following possible temporal framing values: (i) the number of patients per hour in the past day (V-Day), (ii) the average number of patients per hour in the past week (V-Week), and (iii) the average number of patients per hour in the past month (V-Month). Moreover, Kong et al. [13] point out that users elaborate on information more or less deeply to decide whether it is trustworthy based on different situative factors. To avoid introducing some bias, we create each visualization from the same database to keep data trustworthiness equal across all conditions, i.e., the weekly data is a subset of the monthly data since the former reports the last seven days of the latter. Therefore, the visualizations contain all information the user needs to evaluate the quality of the underlying data.

C. Tasks

Our study includes a sequence of decision tasks (Figure 2). Firstly, we focus on the self-calibrated degree of confidence in a taken decision [17] by asking participants to put themselves in a position of having a health emergency and needing to decide to which hospital they should go to (Figure 2, left). Then, we asked each participant to assess the three visualizations simultaneously and to choose between the two hospitals which one they would like to go to using a think-aloud protocol. We prompted subjects to decide solely based on the number of patients that visited the emergency department

(a) The number of patients per hour in the past day.



(b) The average number of patients per hour in the past week.

(c) The average number of patients per hour in the past month.

Fig. 1. Set of visualizations with different time granularity factors.

on the past day, week, or month. We asked the testers to consider that fewer patients would most likely lead to a shorter waiting time. We made up the names of the hospitals, and we randomly assigned the color encoding of each hospital to reduce any potential biases. As we mentioned, we kept the same number of patients in each of the two hospitals between the visualizations. Therefore, we prompt the subject to base their decision on the temporal framing factor. Moreover, we asked each participant to report which of the line charts is their *anchor frame*, i.e., which frame weighted the most on their decision to choose between the two data trends.

Secondly, participants were assigned a random order through which they would interact with each visualization separately (Figure 2, middle). We asked participants to complete three tasks to ensure that they acknowledged the different framing for each visualization. The tasks consisted of (i) finding the number of patients for a specific point in time, (ii) finding what hospital had the greatest growth of patients in a specific time interval, and (iii) choosing one hospital to visit in case of an emergency. After performing the tasks, we invited users to assess the trustworthiness of the visualization with that they interacted.

For the last part, we presented again to each subject the three visualizations simultaneously (Figure 2, right). We then asked participants to choose which hospital they would go to and which visualization weighted the most in their decision, similar to the first part. This repetition of the first part may allow us to understand the self-calibrated degree of confidence in a taken decision [17], i.e., whether performing tasks with each visualization may alter the choice that the participant made at the beginning. By making participants explore the visualizations in more detail, we believe their perceived trustworthiness towards the same visualizations may change. Therefore, the last part of our experience allows us to understand whether an extensive individual analysis of each visualization affects the interaction and decision-making patterns when they must again report the frame they rely on the most to choose between hospitals. In particular, this three-step methodology supports the analysis of whether there is any specific temporal framing that participants trust the most and if that temporal framing functions as an anchor to the overall decision-making.

D. Measures

Demographics We recorded the gender, age, self-reported visual acuity of each participant, and whether they were colorblind by a validated simplified version of the Ishihara test [26]. We also controlled other external factors such as the last time the participant visited a hospital for an emergency (last week, last month, last year, or never), and their familiarity with line charts using a five-point Likert scale ranging from *"Not familiarized"* to *"Completely familiarized"*.

Trust Assessment We assessed visualization trustworthiness with a five-point Likert scale ranging from *"I do not trust this information"* (1) to *"I completely trust this information"* (5) for each visualization framing.

User interaction We measure the **response time** for each tested task in seconds. In addition, we assess the **number of hover events per data point** that the participant triggered while interacting with each graph. A hover event is triggered when the user hovers over a data point to inspect the number of patients through a tooltip.

E. Procedure

We recruited participants through standard convenience sampling procedures by direct contact and word of mouth. Our final data set is composed of 89 participants (38 males, 51 females) between 18 and 69 years old ($M = 27.40; SD = 12.04$). Participants are general end-users with no particular relation to the healthcare area and with normal or corrected-to-normal vision. User tests were conducted through an online videoconference platform, forcing the visualizations to resize to ensure it was displayed in the same physical size, regardless of device resolution. After participants provided informed consent, we first asked them to read a document that introduces the context for the visualizations and prompts participants to be aware of the emergency associated with this crisis while motivating trustworthiness. In addition, the document explains the context of the data of the visualizations and the negative consequences of overcrowding in emergency departments, e.g., the increase in mortality rates.

We continued in a three-part test, as depicted in Figure 2. First, we asked participants to put themselves in a position

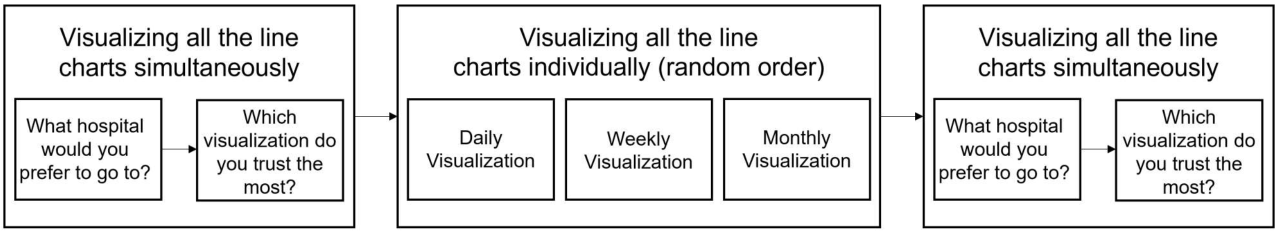


Fig. 2. Overview of the decision processes in our experiment.

of having a health emergency and having the need to decide which hospital they should go to. Participants decided while seeing the three visualizations simultaneously. Then, we asked them to choose which hospital they would prefer to go to and which visualization they trusted the most to make that decision. The assistant collected the time participants took to decide and registered their anchor frame. Next, participants performed the three mentioned tasks separately in random order in each visualization. Additionally, we invited subjects to assess their perceived trust regarding each visualization after interacting with it. The assistant collected the number of hovers that subjects triggered while interacting with each line chart and the time to complete each task. Afterward, we presented to each subject the three visualizations simultaneously. Then, we asked them to choose which hospital they would go to and their anchor frame. Finally, each subject filled in the demographic questionnaire.

F. Research Design and Data Analysis

We ran one-way ANOVAs with the temporal frame (3 levels) as the independent factor to study its effect on the dependent variables (trust perception, decision time, and hovers per point). We also ran two-way mixed ANOVAs with anchor frame (3 levels) and temporal frame (3 levels) as independent factors to understand whether the anchor frame plays a role in these relationships. As we mentioned, the anchor frame is a between-subjects variable and has three possible values: {A-Day, A-Week, A-Month}. The temporal frame is a within-subjects variable with three possible values as well: {V-Day, V-Week, V-Month}. All evaluation sessions were video-recorded to collect interaction metrics. We measured user interaction through two variables: *number of hovers per point*, represented by the sum of hover events that the participant triggered in a visualization divided by the number of data points; and *time to choose a hospital*, which corresponds to the time users take to pick a hospital while analyzing the visualizations individually. We tested for sphericity (Mauchly's test) and used the Greenhouse-Geisser correction when the assumption was not met. ANOVAs were followed by posthoc Tukey's range tests, which include Bonferroni corrections. Finally, we examine whether participants changed their anchor frame between the choice moments. We ran a chi-square test of independence for $r \times c$ contingency tables.

IV. Results

The following subsections discuss the results regarding the self-assessment of trust perception and user interaction metrics. Data are mean $\pm$ standard error unless otherwise stated.

A. Trust Perception

We started studying whether trust perception was influenced by the temporal framing when participants analyzed the different visualizations separately. Results did not show a statistically significant interaction between the framing and perceived trust, $F(1.636, 143.971) = 2.777, p = .076$, partial $\eta^2 = .031$. All distributions look similar with a positive mean rating, yet results suggest that subjects assessed their trust perception with lower grades in V-Week $(4.056 \pm .091)$ compared to V-Day $(4.135 \pm .096)$ and V-Month $(4.213 \pm .088)$ granularity values (Figure 3). In particular, a pairwise comparison reports a statistically significant increase of 0.157 (95% CI, 0.002 to 0.313) points from V-Week to V-Month, $p = .046$.

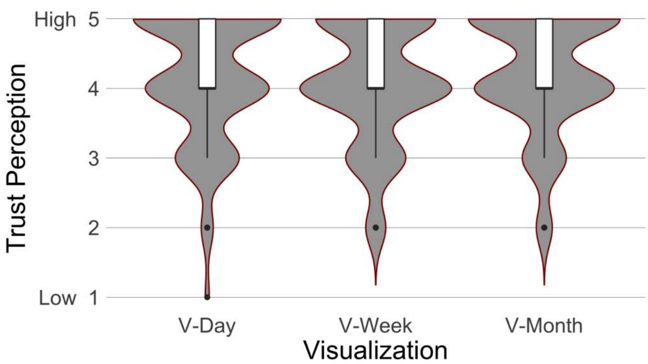


Fig. 3. Violin and boxplots of trust perception for each visualization.

Next, we ran a two-way mixed ANOVA to study whether the anchor frame affects the degree of trust perception per temporal framing (Figure 4). There was a statistically significant interaction between the anchor frame and the temporal framing on trust perception, $F(3.372, 144.982) = 4.469$, $p = .003$, partial $\eta^2 = .094$. In a pairwise comparison, we can observe that, for subjects that chose A-Month, there were statistically significant increases of 0.275 (95% CI, 0.027 to 0.522) points from V-Day to V-Month, $p = .024$, and of 0.333 (95% CI, 0.140 to 0.527) points from V-Week to V-Month, $p < .001$.

These results suggest that people who rely more on A-Month consistently assess their perceived trust in the remaining granularity options (V-Day: $4.020 \pm .127$; V-Week: $3.961 \pm .120$; V-Month: $4.294 \pm .117$). Similarly, people who chose A-Day also trusted more V-Day $(4.417 \pm .261)$ compared to V-Week $(4.250 \pm .248)$ and V-Month $(4.000 \pm .241)$ granularity values. In contrast, people who favored A-Week had similar values across their trust perception for V-Day $(4.231 \pm .178)$, V-Week $(4.154 \pm .169)$, and V-Month $(4.154 \pm .164)$.

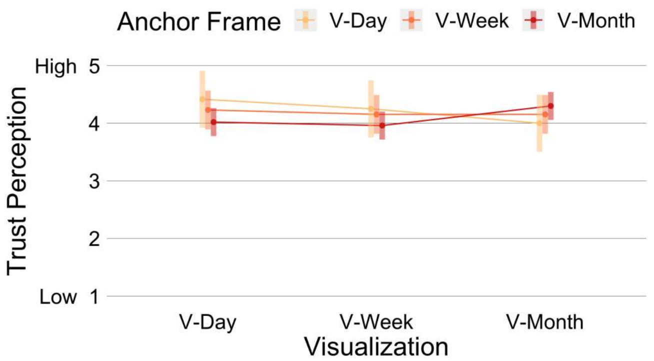


Fig. 4. Estimated marginal means of trust perception based on anchor granularity and the temporal framing of the assessed visualization.

We ran a chi-square test of independence between the anchor frame in both the initial and the final decision moments Table I). There was a statistically significant association between the chosen visualizations, $\chi^2(4) = 60.348, p < .001$. The association was moderately strong [27], Cramer's V $= 0.582$. Overall, these results suggest that participants were consistent in their choices, reflecting a robust self-calibrated degree of confidence in a decision, except for people who chose A-Week since they were more likely to change their decision after a careful analysis of the data. Moreover, A-Month was the most relied-on frame in both decision moments. Our next step is to find whether the interaction data reflects the lower trust perception of the V-Week line chart and how the anchor frame affected how people rated each visualization.

TABLE I
CROSSTABULATION OF THE ANCHOR FRAME IN EACH DECISION MOMENT. ADJUSTED RESIDUALS APPEAR IN PARENTHESES NEXT TO THE OBSERVED FREQUENCIES.

		Final Anchor Frame			
		A-Day	*A-Week*	*A-Month*	Total
Initial Anchor Frame	*A-Day*	8 (5.5)	0 (−1.9)	4 (−2.4)	12
	A-Week	3 (−0.5)	15 (5.4)	8 (−4.2)	26
	A-Month	2 (−3.3)	4 (−3.6)	45 (5.5)	51
	Total	13	19	57	89

B. *User Interaction*

We collected the time in seconds users took to decide between the two hospitals and how many times they hovered a point in the line charts. The temporal framing showed statistically significant effects both in the time users took to decide, $F(2, 174) = 8.221$, $p < .001$, partial $\eta^2 = .086$, as well as in how many hovers per point subjects did, $F(1.457, 126.670) = 23.251$, $p < .001$, partial $\eta^2 = .209$. Regarding the time to decide, participants took less time to choose a hospital when analyzing V-Day $(6.852 \pm .558)$, compared to V-Month $(9.477 \pm .935)$ and V-Week (11.068 ± 1.120) versions. More precisely, a pairwise comparison reported statistically significant increases of 4.216 (95% CI, 1.384 to 7.048) seconds from V-Day to V-Week, $p = .001$, and of 2.625 (95% CI, 0.354 to 4.891) seconds from V-Day to V-Month, $p = .017$. These results suggest that participants find it easier to decide when observing V-Day.

Results showed that participants performed more hover events per points when they analysed V-Week $(0.730 \pm .095)$, followed by V-Day $(.307 \pm .038)$ and then V-Month $(0.297 \pm .042)$. In particular, a pairwise comparison reports statistically significant increases of 0.424 (95% CI, 0.219 to 0.629) hovers per point from V-Day to V-Week, $p < 0.001$, and of 0.434 (95% CI, 0.233 to 0.635) from V-Month to V-Week, $p < 0.001$. Therefore, users interacted more with V-Week, which was the one they rated with a lower trust perception.

Akin to the trust perception analysis, we decided to verify whether the anchor frame affects the time to decide on a hospital and the hovers per temporal framing. We found that the anchor frame significantly affected the time that people took to choose a hospital when they interacted with the visualizations one at a time (Figure 5), $F(3.706, 157.523) = 2.730$, $p = .035$, partial $\eta^2 = .060$. In particular, a pairwise comparison showed that subjects with A-Week had statistically significant increases of 5.385 (95% CI, 1.256 to 9.514) seconds from V-Day to V-Month, $p = .006$. Additionally, subjects that with A-Month showed statistically significant increases of 5.300 (95% CI, 1.540 to 9.060) seconds from V-Day to V-Week, $p = .003$, and of 3.900 (95% CI, 0.676 to 7.124) seconds from V-Month to V-Week, $p = .012$. These results show that people with either A-Day (11.000 ± 3.007) or A-Month (12.680 ± 1.473) take more time to choose a hospital when analyzing V-Week. Nevertheless, only the people who chose A-Day actually take less time to pick a hospital when they analyze their anchor visualization (5.917 ± 1.519). This trend is not present for people with other anchor frame values; those focused on A-Week (6.269 ± 1.032) or A-Month $(7.380 \pm .744)$ are faster in visualization with V-Day.

Finally, we found that the anchor frame did not significantly affect the number of hovers per point performed by the participants when interacting with the visualizations individually (Figure 6), $F(2.864, 123.163) = 0.288$, $p = .825$, partial $\eta^2 = .007$. Even though there was non-significant interaction, a pairwise comparison showed that subjects with an A-Week had a statistically significant increase of 0.437 (95% CI, 0.053 to 0.820) hovers per point from V-Day to V-Week, $p = .020$. Moreover, subjects with A-Month showed statistically significant increases of 0.433 (95% CI, 0.159 to 0.707) hovers per points from V-Day to V-Week, $p = .001$, and of 0.494 (95% CI, 0.226 to 0.761) hovers per points from V-Month to V-Week, $p < .001$.

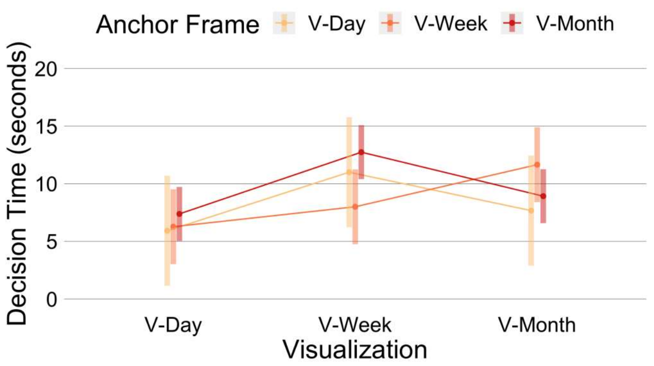


Fig. 5. Estimated marginal means of time to choose a hospital based on the anchor frame and the temporal framing of the assessed visualization.

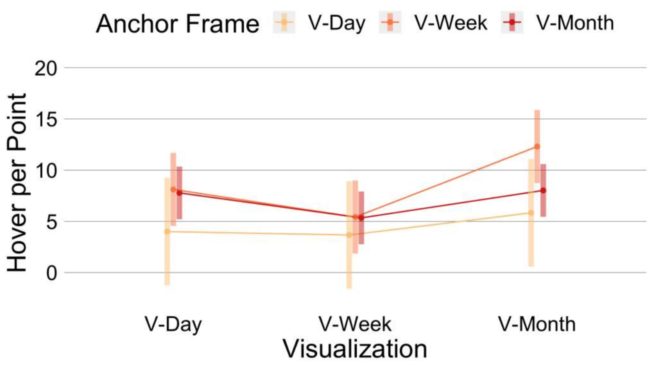


Fig. 6. Estimated marginal means of hovers per points based on the anchor frame and the temporal framing of the assessed visualization.

C. Validity of the Study

We decided to analyze whether any external artifacts affected our study. Regarding user familiarity (Figure 7), we found that most individuals familiarized themselves with line charts (4.47 ± 0.95). Since we asked participants to choose between two data trends in each time frame of the line charts, we need to check whether the decision depends on the hour, weekday, or day of the month when we conducted the test. In particular, we started by analyzing whether the hour, weekday, or day affected the anchor frame. We ran chi-square tests of independence for each time dimension (hour, weekday, day of the month). Results showed that neither the hour ($\chi^2(22) = 15.718, p = .830$), weekday ($\chi^2(10) = 6.196, p = .799$), or day of the month ($\chi^2(36) = 33.056, p = .609$) had an impact on the decision made by the participants.

Afterward, we ran one-way ANOVAs to analyze whether the hour, weekday, or day when participants conducted the study affected the time participants took to choose a hospital and decide which visualization they trusted the most when participants analyzed the visualization for the first time. Again, results showed that neither the hour ($F(5, 83) = 0.443, p = .817$), the weekday ($F(11, 77) = 1.453, p = .167$), or the day of the month ($F(18, 70) = 0.680, p = .819$) had a statically significant impact on the time taken by participants. Then, we

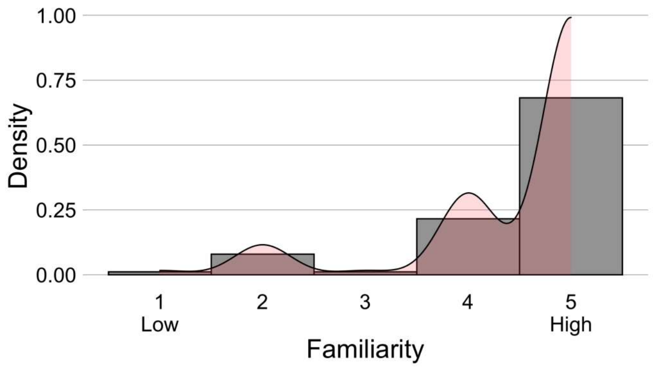


Fig. 7. Histogram with density plot of line chart familiarity.

conducted one-way ANOVAs to analyze the impact of the moment participants conducted the study on the trust perception of each of the framing values. Results showed that neither the hour, the weekday, nor the day of the month had any impact on the trust perception level for each framing visualization. Finally, we verified that the last time each participant visited a hospital did not affect the dependent variables. In particular, the distribution shows that most participants visited a hospital last year (85.39%), and similar amounts to last month (5.62%), week (3.37%), or never (4.49%). These results lead us to believe that the time the subjects conducted our study did not affect their decision-making and trust perception.

V. Discussion

Our results shed new light on the understanding of the effects of temporal framing on trust perception in the context of InfoVis with healthcare data in an emergency.

A. Answering the Research Question

Although there was a non-significant relationship between temporal framing and the trust perceived by the participants, they were more likely to attribute a lower score of perceived trust to V-Week. The high number of hover events per point and higher decision times between hospitals while interacting with V-Week corroborate these results. However, past research in Perceptual Psychology related to subitizing – "the immediate apprehension of the exact number of items in small sets" [28], [29] – suggests that individuals should find it easier to interpret visualizations with a lower number of points. In our case, we expected that individuals would have decreased decision times in V-Week since it had fewer data points and, consequently, individuals could subitize them more easily. In this light, we assume that the individuals actively spent time assessing their trust for each graph, providing some robustness to our results. Moreover, we noticed that while analyzing V-Week, people paid more attention to the peaks than the evolution of the data. It hints toward the data analysis focusing on maximum or minimum values rather than an overall appreciation of the time series.

Contrarily to this trend, results showed that, in general, participants made a decision faster when using V-Day and that they trusted more in V-Month. Regarding V-Day, the short decision time may be an influence of not aggregated (daily) *vs.* aggregated (V-Month and V-Week) values and/or that the concept of patients per hour maps directly to V-Day and, therefore, they make decisions faster. Additionally, participants may have interacted more with V-Week to know more or understand the values behind the aggregation. Contrary to what we expected, our results showed that the number of points presented in a visualization had a more predominant impact on perceived trust and user interactions than the temporal framing of the visualization. This finding is in line with Xiong et al. [10], since participants relied more on a visualization that showed a higher amount of information.

We also analyzed the self-calibrated degree of confidence in a decision by firstly asking the participants to state which temporal frame they would find more reliable in deciding between hospitals. V-Month was the most chosen level when we asked subjects which frame was more relevant to deciding between the two hospitals. Additionally, participants who initially trusted the most in that frame significantly perceived it with higher trust than the remaining framing options. In contrast, participants who initially weighed A-Week the most were more likely to change their decision about which frame weighted the most on their decision after a careful analysis of the data.

B. Additional Findings

We were able to recognize some patterns related to the analysis of time-series visualizations. In general, **participants seem to trust more in visualizations that disclosure information through more data points.** In particular, V-Month displayed 30 points, and it was the one participants trusted the most. This finding was exacerbated through the think-aloud protocol since participants mentioned that they felt more confident predicting future events when analyzing data from a higher period of time. Although V-Month has a higher time range, it is actually the same amount of information, namely the number of patients per hour with any additional aggregation information. Nevertheless, participants may wrongly perceive it as more data points. The fact that participants appeared to trust a visualization with more data points is in line with Xiong et al. [10].

Additionally, we were able to understand that **participants appeared to attribute more relevance to the maximum values when compared to the overall variation of the data when less information is displayed** such as in V-Week (7 points). We noticed for interaction data that **people interact more with a visualization that they trust less**. Researchers may leverage this relationship to adapt the framing of visualization when the system detects a large amount of interaction data since it may indicate that users are not trusting in what they are seeing.

Finally, we repeated the first component of the procedure to reevaluate the change in trust perception from the participant's perspective. As we can observe, there are no significant changes between the decisive moments. Most of the individuals kept the anchor frame after interacting with each one individually. Therefore, we conclude that asking people to rate visualizations explicitly individually does not seem to significantly influence their ratings afterward when all the visualizations are together.

C. Limitations and Future Work

Some relevant factors may explain the lack of significance observed in the effect of temporal framing on perceived trust. The assessment of perceived trust through a Likert scale may have confused participants, as they have assessed their perceived trust quite similarly in each frame. Another possible explanation may be that factual tasks may not have been enough to trigger significant trust variations. Moreover, results showed longer completion times and increased hover events performed when participants analyzed V-Week. Taking a closer look at the V-Week line chart (Figure 1b), the low amount of data points may have led participants to have an exacerbated perception of the peaks. Notice that V-Month (Figure 1c) also has the same peaks in the last seven days, yet the amount of data points reduces the area they cover, hence their reduced impact. Therefore, this design may have led to additional user interaction when participants explored this visualization, so future experiences should use randomly generated datasets. Additionally, future work should do follow-up experiments to understand whether temporal framing, the amount of data shown, or an interaction of the two drive the effect. We can consider different data encodings and a mechanism to differentiate between accidental and intentional hover events. Another limitation is that the V-Day data points are not a subset of the V-Week or V-Month charts. Consequently, the mental model varies since they have to consider new data points compared to the other line charts. These implications are crucial for the experiment and may induce bias in the results.

Future work may leverage temporal aggregation in the visualization design. A general example of temporal aggregations is the use of moving average in the finance domain, where the moving average sums up the data points of financial security over a specific time period and divides the total by the number of data points to obtain an average. In this case, the moving average is continually recalculated based on the latest price data[1]. This approach allows data to be smoother and less unpredictable [30]. Therefore, we could use this approach to study whether the moving average of patients in the different temporal frames affects trust perception.

Future work also includes analyzing perceived trust in light of individual differences such as personality. This type of analysis may explain in-depth relationships between psychological constructs and the trust level of a specific decision. In addition, leveraging a think-aloud protocol to study the use of words, such as "unsure, uncertain, maybe, perhaps", may hold a more accurate assessment of trust compared to the Likert

[1]https://corporatefinanceinstitute.com/resources/knowledge/other/moving-average/

scale assessment. In particular, these words can be indicative of uncertainty or trust, as suggested by Sacha et al. [6]. Finally, it is imperative to mention that uncertainty plays a relevant role in the trust building [6], [7], [31]. Since trust increases when users are aware of the presence of uncertainty in data [6], it is also important to represent this factor while analyzing trust perception.

VI. Conclusions

This study focused on identifying the effects of temporal framing on trust perception in the context of time-oriented visualization. Results show that temporal framing is a relevant feature of time-based visualizations since people trust more in a line chart with more data points and interact more with a line chart that they trust less. The temporal framing that subjects initially rely on the most in the decision tasks also impacts the trust perception when individuals examine visualizations one at a time. These contributions are relevant for designers of visual analytic systems, particularly when studying human decision-making supported by visualization. They provide implications to understand trust perception in a health emergency scenario and its variation inline chart techniques differing in the time-based granularity.

Acknowledgment

This work was supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with references SFRH/BD/144798/2019 and UIDB/50021/2020.

References

[1] M. Card, *Readings in information visualization: using vision to think*. Morgan Kaufmann, 1999.

[2] Y. Zhang, R. K. Bellamy, and W. A. Kellogg, "Designing information for remediating cognitive biases in decision-making," in *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, 2015, pp. 2211–2220.

[3] E. Dimara, G. Bailly, A. Bezerianos, and S. Franconeri, "Mitigating the attraction effect with visualizations," *IEEE transactions on visualization and computer graphics*, vol. 25, no. 1, pp. 850–860, 2018.

[4] E. Dimara, S. Franconeri, C. Plaisant, A. Bezerianos, and P. Dragicevic, "A task-based taxonomy of cognitive biases for information visualization," *IEEE transactions on visualization and computer graphics*, vol. 26, no. 2, pp. 1413–1432, 2018.

[5] A. Kale, M. Kay, and J. Hullman, "Visual reasoning strategies for effect size judgments and decisions," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 2, pp. 272–282, 2020.

[6] D. Sacha, H. Senaratne, B. C. Kwon, G. Ellis, and D. A. Keim, "The role of uncertainty, awareness, and trust in visual analytics," *IEEE transactions on visualization and computer graphics*, vol. 22, no. 1, pp. 240–249, 2015.

[7] E. Mayr, N. Hynek, S. Salisu, and F. Windhager, "Trust in information visualization," in *EuroVis Workshop on Trustworthy Visualization (TrustVis), Robert Kosara, Kai Lawonn, Lars Linsen, and Noeska Smit (Eds.). The Eurographics Association. DOI: http://dx. doi. org/10.2312/trvis*, vol. 20191187, 2019.

[8] J. Thomas and J. Kielman, "Challenges for visual analytics," *Information Visualization*, vol. 8, no. 4, pp. 309–314, 2009.

[9] R. Salway, R. Valenzuela, J. Shoenberger, W. Mallon, and A. Viccellio, "Emergency department (ed) overcrowding: evidence-based answers to frequently asked questions," *Revista Médica Clínica Las Condes*, vol. 28, no. 2, pp. 213–219, 2017.

[10] C. Xiong, L. Padilla, K. Grayson, and S. Franconeri, "Examining the components of trust in map-based visualizations," 2019.

[11] A. K. Schnackenberg and E. C. Tomlinson, "Organizational transparency: A new perspective on managing trust in organization-stakeholder relationships," *Journal of Management*, vol. 42, no. 7, pp. 1784–1810, 2016.

[12] E. Costante, J. Den Hartog, and M. Petkovic, "On-line trust perception: What really matters," in *2011 1st Workshop on Socio-Technical Aspects in Security and Trust (STAST)*. IEEE, 2011, pp. 52–59.

[13] H.-K. Kong, Z. Liu, and K. Karahalios, "Trust and recall of information across varying degrees of title-visualization misalignment," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–13.

[14] L. Bartram, M. Correll, and M. Tory, "Untidy data: The unreasonable effectiveness of tables," *arXiv preprint arXiv:2106.15005*, 2021.

[15] A. Dasgupta, S. Burrows, K. Han, and P. J. Rasch, "Empirical analysis of the subjective impressions and objective measures of domain scientists' visual analytic judgments," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017, pp. 1193–1204.

[16] K. Kelton, K. R. Fleischmann, and W. A. Wallace, "Trust in digital information," *Journal of the American Society for Information Science and Technology*, vol. 59, no. 3, pp. 363–374, 2008.

[17] A. Dasgupta, J.-Y. Lee, R. Wilson, R. A. Lafrance, N. Cramer, K. Cook, and S. Payne, "Familiarity vs trust: A comparative study of domain scientists' trust in visual analytics and conventional analysis methods," *IEEE transactions on visualization and computer graphics*, vol. 23, no. 1, pp. 271–280, 2016.

[18] D. Sprague and M. Tory, "Exploring how and why people use visualizations in casual contexts: Modeling user goals and regulated motivations," *Information Visualization*, vol. 11, no. 2, pp. 106–123, 2012.

[19] M. M. Roghanizad and D. J. Neufeld, "Intuition, risk, and the formation of online trust," *Computers in Human Behavior*, vol. 50, pp. 489–498, 2015.

[20] W. Aigner, S. Miksch, W. Müller, H. Schumann, and C. Tominski, "Visual methods for analyzing time-oriented data," *IEEE transactions on visualization and computer graphics*, vol. 14, no. 1, pp. 47–60, 2007.

[21] M. Adnan, M. Just, and L. Baillie, "Investigating time series visualisations to improve the user experience," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2016, pp. 5444–5455.

[22] M. Waldner, A. Diehl, D. Gračanin, R. Splechtna, C. Delrieux, and K. Matković, "A comparison of radial and linear charts for visualizing daily patterns," *IEEE transactions on visualization and computer graphics*, vol. 26, no. 1, pp. 1033–1042, 2019.

[23] N. Oscar, S. Mejía, R. Metoyer, and K. Hooker, "Towards personalized visualization: Information granularity, situation, and personality," in *Proceedings of the 2017 Conference on Designing Interactive Systems*, 2017, pp. 811–819.

[24] W. Javed, B. McDonnel, and N. Elmqvist, "Graphical perception of multiple time series," *IEEE transactions on visualization and computer graphics*, vol. 16, no. 6, pp. 927–934, 2010.

[25] J. Heer, N. Kong, and M. Agrawala, "Sizing the horizon: the effects of chart size and layering on the graphical perception of time series visualizations," in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2009, pp. 1303–1312.

[26] D. V. de Alwis and C. H. Kon, "A new way to use the ishihara test," *Journal of neurology*, vol. 239, no. 8, pp. 451–454, 1992.

[27] J. Cohen, *Statistical power analysis for the behavioral sciences*. Academic press, 2013.

[28] H. Railo, M. Koivisto, A. Revonsuo, and M. M. Hannula, "The role of attention in subitizing," *Cognition*, vol. 107, no. 1, pp. 82–104, 2008.

[29] M. Piazza, A. Fumarola, A. Chinello, and D. Melcher, "Subitizing reflects visuo-spatial object individuation capacity," *Cognition*, vol. 121, no. 1, pp. 147–153, 2011.

[30] A. Raudys, V. Lenčiauskas, and E. Malčius, "Moving averages for financial data smoothing," in *International Conference on Information and Software Technologies*. Springer, 2013, pp. 34–45.

[31] J. Zhou, S. Luo, and F. Chen, "Effects of personality traits on user trust in human–machine collaborations," *Journal on Multimodal User Interfaces*, pp. 1–14, 2020.