Virtual Spaces, Real Perceptions: Analyzing Crowding in Virtual Environments

by

Gabrielle Aparecida Pires Alves
M.B.A., University of Sao Paulo, 2018
M.Sc., Universy of Sao Paulo, 2023

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Computer Science

Virtual Spaces, Real Perceptions: Analyzing Crowding in Virtual Environments

by

Gabrielle Aparecida Pires Alves
M.B.A., University of Sao Paulo, 2018
M.Sc., Universy of Sao Paulo, 2023

Supervisory Committee

Dr. Prof. Brandon Haworth, Supervisor
(Department of Computer Science)

Dr. Prof. Yvonne Coady, Departmental Member
(Department of Computer Science)

ABSTRACT

This thesis explores the dynamics of crowding perceptions and emotional responses within virtual environments, focusing on the impact of avatar density and representation. The study aims to extend the understanding of how digital spaces influence user experiences, particularly as interactions increasingly transition to virtual platforms.

The research problem centers on how varying crowd densities and different avatar representations affect user perceptions and emotional responses. The study addresses four primary research questions: the consistency of crowding perceptions and emotional responses (RQ1), the threshold densities at which emotional responses significantly change (RQ2), the impact of different avatar representations on perceptions and responses (RQ3), and the linear relationship between avatar density and perceived crowding (RQ4).

Using a within-subjects experimental design, participants were exposed to virtual environments with varying densities of particle, 2D, and 3D avatars. Data were collected on users' crowding perceptions and emotional responses (valence, arousal, dominance). Results revealed significant inconsistencies in crowding perceptions and emotional responses across multiple exposures, highlighting the critical role of avatar representation. 3D avatars led to more consistent experiences compared to particle and 2D avatars. Threshold densities at which emotional responses changed were identified, with density 8 being a significant point across various measures. Different avatar representations significantly impacted perceptions and responses, with particle avatars generally resulting in less intense emotional responses. Linear relationships were found between avatar density, perceived crowding, and emotional responses.

The findings emphasize the importance of avatar representation in virtual environment design, contributing to a broader understanding of crowding in both virtual and physical settings. This research bridges the gap between traditional architectural practice and virtual environments, offering practical recommendations for enhancing user comfort, satisfaction, and well-being in digital spaces.

Contents

List of Tables

List of Figures

ACKNOWLEDGEMENTS

I would like to thank:

Brandon Haworth, during one of the most vulnerable times in my life, your compassion and kindness became a lifeline I desperately needed. You offered more than just guidance; you provided unwavering support, endless encouragement, and the patience to see me through every step of this journey. While you are undoubtedly an exceptional mentor, it's your humanity, your genuine care and warmth, that truly set you apart. Thank you!

My parents, who sacrificed so much and poured endless love into every moment of my life.

My family, who have been my best friends throughout my entire life. I miss you all every single day.

My friends, both those who have stood by me for many years and those I've met along this master's journey. And to those who, in the final stretch, offered their unwavering support and even a place to stay, helping me through the most intense part of writing this thesis—I am deeply grateful.

Livros & Citações, what began as a project in 2010 has grown alongside me. Thank you to the dear followers and friends who have supported my craziness and extended a hand when I needed it most. As I close another chapter of my life, I owe so much of it to all of you.

"Daughter, your faith has healed you. Go in peace and be freed from your suffering."
Mark 5:34

"I've still got joy in chaos.
I've got peace that makes no sense.
I won't be going under.
I'm not held by my own strength."
Firm Foundation (He Won't) by Maverick City Music

DEDICATION

Over these past two years, I broke and rebuilt myself, learning lessons I never wished to face. I learned to manage panic attacks, discovered that there's no shame in being vulnerable, nor in giving up and choosing a new path. I also discovered myself—my strength, my potential, and my worth. I realized that it becomes easier to move forward when you remember where you came from, because the worst that can happen is returning home, and my home is my favorite place in the world. So, I dedicate this thesis to myself, to the girl I once was, to all the dreams I had, the dreams I fulfilled, and to the new dreams that lie ahead.

Finally, I dedicate this to my home. To my parents, who never had the opportunity to finish their education, yet are my greatest mentors in perseverance and hard work, and my deepest source of pride. To my siblings, each so different, yet together we form something perfect. To my sister-in-law, the long-lost sister we found along life's journey. To my nephews, the greatest loves of my life. To my aunt, my second mother. To my grandmother, a true fortress. To the friends I met along this journey who offered comfort and care. And to Shayane, who, despite the distance, was always the closest. *And to the fifteen-year-old girl who once dreamed—you can stop dreaming now, we made it.*

Chapter 1

Introduction

No matter what happens in the world of human beings, it happens in a spatial setting, and the design of that setting has a deep and persisting influence on the people in that setting.

The Hidden Dimension,
E. T. Hall [49]

The way people perceive crowding greatly affects their experiences in the spatial setting, including public transit, shopping centers, and events [44, 20]. Our sense of the density of people around us influences how comfortable, anxious, or pleased we feel [23]. This perception can significantly differ depending on the context. For example, festivals frequently elicit happy emotions despite the presence of large crowds, yet the same degree of crowding in public transportation might cause discomfort and stress [17, 51].

The pollution, crime, and social instability worries in metropolitan regions in the 1960s and 1970s made crowding research increasingly popular [78, 65]. Evidence from this time period suggests that overpopulation negatively affects people's mental health and their ability to engage with others and the natural world [21]. This has led to a growing reliance on understanding how people experience crowding in many fields, including transportation, urban planning, engineering, psychology, and business [85, 67, 47, 19].

Extensive research, both theoretical and practical, has explored the concept of crowding in the field of architecture [72, 60, 15]. Researchers are investigating the impact of crowding on people's desires and preferences in different situations, such as workplaces, schools, and stadiums, with the aim of optimizing space utilization and social density [18, 52, 50].

Yet, conventional approaches to design have not undergone much change, even if our understanding of how people behave in congested environments has progressed [71]. Architects often employ orthographic slices and other two-dimensional representations to depict environments. The two-dimensional (2D) floorplans are great for seeing potential layouts, but they don't show you the genuine feel and functionality of the space. This disparity further complicates the connection between visual communication and architecture, highlighting the necessity for increasingly advanced technologies such as three-dimensional (3D) modeling, virtual reality, and augmented reality [38, 36]. These cutting-edge visualization methods provide interactive and immersive means of efficiently designing and communicating ideas [33].

This issue is particularly evident in residential buildings. It is crucial to thoroughly analyze the diverse living experiences and preferences of occupants [53]. Enhanced design strategies that are more true to life and engaging are necessary, as 2D depictions often fall short in capturing this level of complexity. Architects can utilize advanced techniques to create places that meet the unique needs of their customers.

While tools like 3D modeling and AutoCAD are acknowledged, their integration into current design processes remains limited in addressing human factors adequately [71]. Architects are often trained to understand and create 2D floorplans as quick alternatives for space visualization. However, this approach does not fully consider the complexities of human interaction with space, including social distance and neurodiversity.

Parallel to these considerations in physical environments, my research extends this understanding to virtual environments. I examined how different avatar types and densities affect users' experiences in digital spaces. Similar to physical environments, virtual crowding can influence users' comfort and emotional states, which is increasingly important as more interactions shift to digital platforms.

My goal in using particle-based, 2D, and 3D avatar representations was to study how these visual cues affect users' sense of space, presence, and emotional well-being in virtual spaces. I used a within-subjects strategy to gather data from people who were recruited in various ways. Visual representations of areas with varying degrees of crowding and avatar types were reviewed by participants, ensuring robust and unbiased conclusions. In order to determine how crowd density and avatar representation affect the sensation of crowding, our study collected a large amount of data. The results will provide insights into creating both digital and real environments.

In summary, this thesis explores the impact of avatar density and representation in virtual environments on user perceptions and emotional reactions, deepening our understanding of crowding perception. By carefully studying several avatar types at varying crowd densities, this research offers profound insights into the emotional and psychological dynamics of virtual crowding. These findings are crucial for fostering more enjoyable experiences, greater engagement, and improved social interactions in digital contexts.

This thesis bridges the gap between conventional architectural practice and virtual settings by offering new perspectives on collaborative space design. Using virtual crowding as a tool, it investigates the effects of various spatial configurations and crowd levels on human behavior and mental health. Applying this knowledge can result in sustainable design processes that create environments supportive of efficiency and mental health, ultimately enhancing overall enjoyment and well-being. Additionally, the relevance of visualization in efficiently communicating urban planning concepts to various stakeholders and the utility of collaborative design in virtual environments for urban planning is emphasized, allowing numerous users to participate and provide input.

1.1 Motivation and Research Questions

Traditionally, architects have used 2D, flat drawings to represent their projects. These orthographic slices provide a bird's-eye view of spatial arrangements, but they fall short of conveying how a space will truly feel and function. While architects can mentally translate these 2D plans into 3D spaces, the general public sometimes has trouble doing the same. This gap in understanding emphasizes the need for more sophisticated tools—like 3D modeling and virtual reality—to bring designs to life in ways everyone can grasp.

Now, fast forward to the digital age. Virtual environments are becoming increasingly important, not just for gaming or entertainment but for work, social interactions, and even therapy. Here, the concept of crowding takes on new dimensions. Just as in physical spaces, virtual crowding can influence comfort and emotional states. But how do these digital crowds—represented by avatars—affect our perceptions and feelings?

To unravel this, I set out on a research journey. I started by investigating how consistent people's emotional reactions and perceptions of crowding are. Just picture yourself in a virtual area populated by avatars. Is your emotional reaction and perception of congestion consistent, or does it alter every time you visit? When creating trustworthy and predictable virtual worlds, consistency is king. All users should have the same consistent experience in virtual places, and this can be achieved if we comprehend this. In light of this, the first research question is as follows:

RQ1: How consistent are users' crowding perceptions and emotional responses (valence, arousal, and dominance) when exposed to varying crowd densities and different avatar representations (particle, 2D, and 3D) multiple times?

From this research question, four hypotheses have been generated, one of which is the null hypothesis, and will be examined in this work:

$RQ1.H_0$**:** If users are exposed to varying crowd densities with different avatar representations (particle, 2D, and 3D), their crowding perceptions and emotional responses (valence, arousal, and dominance) will remain consistent across multiple exposures.

$RQ1.H_1$**:** If users are exposed to varying crowd densities with particle avatar representations, their crowding perceptions and emotional responses (valence, arousal, and dominance) will significantly differ across multiple exposures.

$RQ1.H_2$**:** If users are exposed to varying crowd densities with 2D avatar representations, their crowding perceptions and emotional responses (valence, arousal, and dominance) will significantly differ across multiple exposures.

$RQ1.H_3$**:** If users are exposed to varying crowd densities with 3D avatar representations, their crowding perceptions and emotional responses (valence, arousal, and dominance) will significantly differ across multiple exposures.

Additionally, I wanted to determine the density at which users' emotional responses begin to alter significantly. The literature suggests that when density exceeds a certain level, people experience discomfort due to crowding. Like in the physical world, virtual places probably have a certain density, beyond which being there becomes overwhelming. Identifying this threshold aids in the design of surroundings that minimize negative emotional affects, resulting in a pleasant experience for all users. That brought us to our second research question:

RQ2: Is there a threshold density at which users' emotional responses significantly change?

From this research question, four hypotheses have been generated, one of which is the null hypothesis, and will be examined in this work:

$RQ2.H_0$**:** If avatars are represented in different formats, the thresholds for significant changes in users' emotional responses will remain the same regardless of the representation (particle, 2D, or 3D).

$RQ2.H_1$**:** If particle representations of avatars are used, the threshold density for significant changes in users' emotional responses will be higher compared to 2D avatar representations.

$RQ2.H_2$**:** If particle representations of avatars are used, the threshold density for significant changes in users' emotional responses will be higher compared to 3D avatar representations.

$RQ2.H_3$**:** If 2D avatars are used, the threshold density for significant changes in users' emotional responses will be higher compared to 3D avatar representations.

The next step was for me to learn how various avatar types—such as particles, 2D figures, and intricate 3D models—impact users' emotional reactions and perceptions of crowding. Consider what it would be like to walk into a room full of simple dots instead of realistic 3D avatars. Your perception of crowd size and your emotional response to it can be significantly changed by the way these avatars are portrayed. Having this knowledge is crucial for creating virtual settings that are useful for both practical and emotional reasons. As a result, the third research question was born:

RQ3: Do different types of avatar representations (particle, 2D, and 3D) and varying densities of avatars in a virtual environment lead to differences in users' perceptions of crowding and their emotional responses (valence, arousal, and dominance)?

From this research question, four hypotheses have been generated, one of which is the null hypothesis, and will be examined in this work:

$RQ3.H_0$**:** If different types of avatar representations (particle, 2D, and 3D) are placed in an environment, there will be no significant differences in users' perceptions of crowding and their emotional responses (valence, arousal, and dominance).

$RQ3.H_1$**:** If particle representations of avatars are placed in an environment, users' perceptions of crowding and their emotional responses (valence, arousal, and dominance) will be less intense compared to when 2D avatars are used.

$RQ3.H_2$**:** If particle representations of avatars are placed in an environment, users' perceptions of crowding and their emotional responses (valence, arousal, and dominance) will be less intense compared to when 3D avatars are used.

$RQ3.H_3$**:** If 2D avatars are placed in an environment, users' perceptions of crowding and their emotional responses (valence, arousal, and dominance) will be less intense compared to when 3D avatars are used.

Finally, I explored the relationship between avatar density and perceived crowding. Is this relationship linear, or does it follow a more complex pattern? Understanding this helps in predicting how changes in avatar numbers will affect users' experiences and emotions. This insight is vital for managing virtual spaces, whether for virtual meetings, social gatherings, or online marketplaces. This brought forth the fourth research question:

RQ4: Is there a linear relationship between avatar density and perceived crowding, and how do emotions (valence, arousal, dominance) vary linearly with avatar density for different types of avatar representations?

From this research question, seven hypotheses have been generated, one of which is the null hypothesis, and will be examined in this work:

$RQ4.H_0$**:** If avatar density increases, there will be no linear relationship between avatar density and perceived crowding, nor will there be a linear variation in emotional responses (valence, arousal, and dominance) for different types of avatar representations (particle, 2D, and 3D).

$RQ4.H_1$**:** If avatar density increases, there will be a linear relationship where perceived crowding increases for particle representations of avatars.

$RQ4.H_2$**:** If avatar density increases, there will be a linear relationship where perceived crowding increases for 2D representations of avatars.

$RQ4.H_3$**:** If avatar density increases, there will be a linear relationship where perceived crowding increases for 3D representations of avatars.

$RQ4.H_4$**:** If avatar density increases, the emotional responses (valence, arousal, and dominance) will vary linearly for particle representations of avatars.

$RQ4.H_5$**:** If avatar density increases, the emotional responses (valence, arousal, and dominance) will vary linearly for 2D representations of avatars.

$RQ4.H_6$**:** If avatar density increases, the emotional responses (valence, arousal, and dominance) will vary linearly for 3D representations of avatars.

Through this journey, I aim to bridge the gap between traditional architectural design and the evolving needs of virtual environments. By investigating how avatar representations and densities affect our perceptions and emotions, I can help create virtual and physical spaces that truly resonate with our psychological and emotional needs. My ultimate goal is to enhance comfort, satisfaction, and well-being, whether navigating a virtual event or moving through a physical building.

This research not only deepens our understanding of crowding in virtual environments but also offers valuable insights for real-world design. By leveraging these findings, architects and designers can create more inclusive, emotionally supportive, and user-friendly spaces. In essence, this study is about making environments—both digital and physical—that feel just right for everyone.

By incorporating avatars into the design process, architects can create more user-centered environments. Realistic avatars can reveal subtle yet significant issues that might not be apparent through traditional design methods. For example, they can highlight areas that may feel cramped or uncomfortable, suggest better placements for amenities, and improve wayfinding solutions.

Moreover, the use of avatars can facilitate communication between designers, stakeholders, and clients by providing a more tangible and relatable representation of the human experience within the proposed designs.

1.2 Research Objectives

The main objective of this thesis is to examine how different types and quantities of avatars impact our comfort, emotions, and overall experiences in virtual environments. The study explores the impact of different types and numbers of avatars on our emotions and interactions in virtual environments, whether they are simple dots, flat shapes or realistic 3D individuals.

Many times, traditional architecture design employs difficult-to-understand, flat, two-dimensional drawings. Using particle representations of people and 3D modeling, this work takes one step further in trying to close the gap between outdated techniques and the real requirements and perceptions of users.

Picture yourself stepping into a virtual room filled with lively avatars. How do the number of these digital figures affect how you perceive space and your overall state of mind? How does the visual appearance of different elements impact your emotions? Whether they are simple dots, flat shapes, or intricate 3D figures, how does it influence your feelings? This research aims to address these questions, providing insights that can assist designers in developing environments—whether virtual or physical—that are not only more efficient and user-friendly, but also more emotionally supportive and inclusive.

The ultimate goal is to change how we think about space and design, making environments that are more in tune with our psychological and emotional needs. By understanding how we perceive crowding and how it affects our emotions, this study aims to create spaces that enhance our comfort and satisfaction, whether we are navigating a virtual event or moving through a physical building.

1.3 Thesis Overview

In answering the above RQs, this work provides the following contributions:

Contribution 1: We conducted an in-depth analysis of users' perceptions of crowding and their emotional responses (valence, arousal, and dominance) in virtual environments with varying densities of avatars. Our findings reveal significant differences in crowding perceptions and emotional responses based on the type of avatar representation (particle, 2D, and 3D). This highlights the importance of avatar representation in virtual environment design.

Contribution 2: Our study identified threshold densities at which users' emotional responses significantly change. This insight can inform the design of virtual environments to optimize user experience and emotional well-being. .

Contribution 3: We investigated the consistency of users' crowding perceptions and emotional responses when exposed to varying crowd densities and different avatar representations multiple times. Our results show that crowding perceptions and emotional responses are not consistent across repeated exposures, varying significantly based on the type of avatar representation. This underscores the need for careful consideration of avatar representation for collaborative design.

Chapter 1 introduces a summary of the topic, including definitions and the research questions. It also presents the open problem that will be addressed, along with its context, impact, and overall motivation for the research.

Chapter 2 provides an overview of related work in the field of crowding perception and architectural design. It discusses previous research studies and highlights gaps in the existing literature that this thesis aims to address. The review serves as a foundation for the thesis's research questions and methodology. Specific focus areas include a psychological theory of crowding and the differences between 2D, 3D, and particle representations.

Chapter 3 describes the research methodologies used in this thesis, including quantitative data collection and analysis techniques. It outlines the study design, including the densities and avatar representations used in the experiments, and the survey used for measuring crowding perception. The chapter also addresses each research question individually, explaining the approaches taken to ensure data quality and ethical compliance.

Chapter 4 presents the findings of the thesis. It provides a detailed analysis of the study execution and participant responses. This chapter answers all the research questions, analyzing user consistency and perception for repeated stimuli, and exploring the influence of avatar representations and densities on crowding perception and emotional responses. The analysis includes various statistical measures and comparisons to ensure robustness.

Chapter 5 reiterates the research findings and discusses their implications. The discussion highlights the practical implications for both virtual environment designers and architectural stakeholders, considering impacts on the design of physical or digital environments.

Chapter 6 summarizes the key findings and discussions of the thesis. The chapter consists of a restatement of the claims and results of the thesis and discusses potential future work. It provides a critical evaluation of the methodology, limitations, and potential threats to validity. It also provides a final reflection on the study's impact, contributions, and suggests future research directions based on the limitations of the current thesis.

Chapter 2

Literature Review

2.1 Introduction

This chapter begins by reviewing crowding perception as a whole, providing a brief history and theoretical background, to contextualize this work. Following this, a review of how density and avatars are represented in architectural designs is provided. Finally, I narrow down to a more specific scope, concluding this chapter with related works and a brief discussion on how this work is novel compared to them.

My work combines psychological theory, computer science, and architectural design, creating an interdisciplinary study that aims to unravel the complexity behind crowding perception. The value of this integrated approach arises not only from its emphasis on interdisciplinarity, but also from the requirement for a broad perspective when investigating human-environment interactions in both the real and digital worlds.

2.2 A psychological theory of crowding

This section explores the theoretical foundations upon which my thesis is based, examining a wide range of concepts related to crowding perception. It investigates how individuals experience and interpret crowded environments.

2.2.1 Density versus Crowding

The presence of crowding may manifest across several socioeconomic layers, ranging from packed residences to large urban centers, which poses challenges in comprehending its nature. An important obstacle is the absence of consistency on the term "crowding," which is often confused with "density" [12, 43].

Stokols [77] clarified the distinction by stating that density is a measure of people or objects per unit area, while crowding is a motivational state resulting from perceived space limitations. According to Stokols, crowding may occur when an individual desires more space than is available. Density alone cannot explain crowding because personal and social factors significantly influence it. Therefore, researchers must consider these factors when studying crowding. By examining crowding perception as countable, measurable phenomena and its spatial, social, and personal aspects, we can better manage it. This thesis focuses on Stokols' definition of crowding, emphasizing the importance of distinguishing it from density.

Perceived crowding due to physical density and individual space requirements can worsen discomfort, stress, and negative emotional responses, such as anxiety and irritability, especially among those intolerant of close proximity. Stokols' model in Figure 2.1 illustrates how humans respond to crowding [78].

2.2.2 Exploring the Complexities of Crowding

Crowding perception affects how we deal with different environments, from crowded streets during peak hours to calm and peaceful places in nature. Over time, as the state-of-the-art evolved, various frameworks and taxonomies were created to define and categorize crowding perception [22, 51, 6]. Among them is the division of crowding perception into two categories: social crowding and non-social crowding [78].

In urban areas or crowded settings like events or concerts, the term "social crowding" refers to the phenomenon whereby the presence of others alters our perception. For instance, during rush hour after work, many of us have to compete for a spot on the bus or move through a crowd of people to walk. According to Sun and Budruk [79], how our culture shapes who we are as individuals also affects how we deal with social crowding.

Cultural anthropologist Edward T. Hall studied how culture affects how we use space and adjust social distances. He popularized the term "proxemics" [49], the study of how individuals from various cultural backgrounds understand and utilize social and personal space. According to Hall's theory, our comfort levels with crowding and interpersonal distances are not the same for everyone and can differ greatly based on cultural norms and customs.

Imagine you are a boy who grew up in a community in Brazil called a favela. You grew up in a place known for its cramped infrastructure, where it's common to hear your neighbor's music playing in your room and have several siblings living with you, getting used to sharing spaces or walking very close to others in narrow streets. This experience shapes you to adapt to high densities of people and resources, where noise or proximity to others may not affect you negatively as it might affect those with different experiences and cultures. However, in some cultures, close physical contact can be uncomfortable and unwanted, and personal space is highly valued, with even slight intrusions causing discomfort.

Environmental psychologist Robert Sommer, in "Personal Space: The Behavioral Basis of Design [76]," examines how the type of activity being conducted in a particular environment affects crowding perception as well. A busy marketplace might feel less crowded than a cramped elevator, even if both have the same number of people, because of the nature of activities and expectations in those spaces [28]. In summary, there are no neutral environments. Humans develop environments, and once constructed, these environments shape the humans who inhabit them. After all, every environment generates physical and psychological reactions in humans.

These two researchers' reflections on how we, as culturally formed individuals, fit into environments with particular objectives and activities highlight the complexity and multitude of factors that influence how people perceive crowding.

The term "non-social crowding", on the other hand, arises from physical factors without the direct presence of other people. People may feel crowded even when they are alone if they don't have enough personal space or privacy. Also, a cluttered or unorganized setting can lead to feelings of non-social crowding, as well as loud noises and bright lights [55]. This type of crowding can be experienced in natural environments or built spaces where the layout or conditions create a sense of confinement. For instance, going for a hike in Canada on a narrow and winding trail in a dense forest affects us differently than being in a cluttered room filled with furniture.

From these two categories of crowding, we can direct the next phase of our background discussion related to the literature on crowding perception towards how it is influenced by both environmental and personal factors [16, 66]. Personal factors include an individual's emotional state, personality, and sense of control over their environment. Someone who feels stressed or anxious is likely to perceive a crowded space as more overwhelming than someone who feels calm and in control. Environmental factors such as the amount and arrangement of space, temperature, and noise levels play crucial roles in non-social crowding. A small room with loud machinery might feel oppressive, even if you are alone. The arrangement of physical space can thus significantly impact crowding perception.

Group dynamics also play a role. In social settings, the type of activity and the relationships among individuals can affect crowding perceptions [26]. A gathering of close friends might feel less crowded than a similar-sized group of strangers because of the comfort and familiarity among friends.

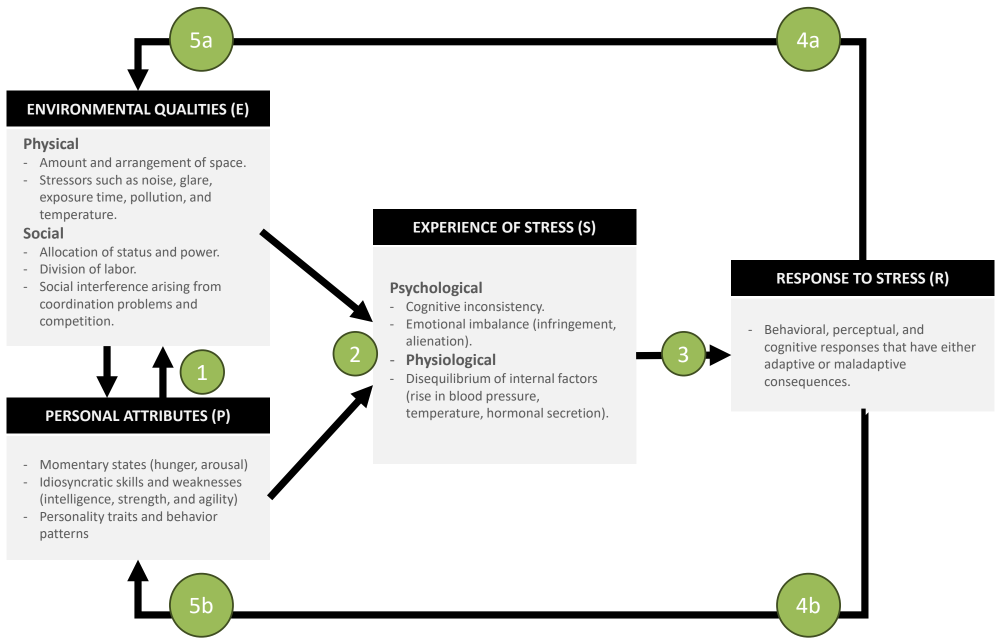

Figure 2.1: Human response to crowding from Stolko, 1972. The model is dividing crowding perception through four basic dimensions. These include: personal attributes (P) such as temporary conditions, individual differences, long-term personality traits, and habitual behaviors; environmental qualities (E), covering physical factors like space and noise, and social factors like allocation of status and division of labor; the experience of stress (S), involving psychological elements like cognitive inconsistencies and emotional imbalances, and physiological elements like internal disequilibrium; and the response to stress (R), which encompasses the behavioral, perceptual, and cognitive responses that can be adaptive or maladaptive based on the individual's coping mechanisms.

A model create by Stokols [78] helps to understand how these different factors navigate between different dimensions. As illustrated in Figure 2.1, the model is divided into four basic dimensions in crowding perception, which range from personal attributes (P), including temporary conditions, individual differences, long-term personality traits, and habitual behaviors that influence crowding perception; to environmental qualities (E), which encompass environmental factors, social structures, and interactions that affect perception; to the experience (S) resulting from this, including cognitive inconsistencies, emotional imbalances, and internal physiological imbalances caused by stress; and the response (R) to that experience, reflecting the various behavioral, perceptual, and cognitive responses individuals exhibit when faced with stress. These responses can have adaptive (positive) or maladaptive (negative) consequences depending on the situation and the individual's coping mechanisms.

There are five main steps in how people experience and deal with crowding, according to the model. These steps happen in order:

(1) Interaction: How the environment (E) and personal attributes (P) affect a person. This determines what they notice and how important it is to them.

(2) Perception: How the interaction between the environment and personal attributes makes someone feel crowded and stressed, both mentally and physically.

(3) Response: How a person tries to reduce their stress, through changes in thinking, perception, or behavior.

(4) Action: The person does something specific to cope with the crowding, like moving to a less crowded area (environmental change, 4a) or changing how they view the situation to make it more bearable (personal change, 4b).

(5) Outcome: The results of these changes, which can be positive (adaptive) or negative (maladaptive), affecting both the environment (5a) and the person (5b).

In the present framework, a response is adaptive to the extent that it relieves either environmental or personal sources of strain and breaks the cycle of crowding stress. A response is maladaptive to the degree that it intensifies stress due to environmental and personal factors and thereby perpetuates the cycle of crowding stress. According to this model, there is a complex interplay between individual characteristics, stressful experiences, and environmental factors that affect stress responses. Each component can interact with others, shaping the overall response to crowding perception in adaptive or maladaptive ways. This framework helps us gain a better psychological understanding of crowding that can help us manage its social impacts and broader consequences.

2.3 Architecture, Crowd Simulation and Avatar Representations

Architecture plays a crucial role in shaping human experiences and behaviors within a built environment. The purpose of any building, whether for residence, leisure, or work, is to support human life [84].

The design of spaces can influence how people interact, move, and feel. Understanding the impact of architectural elements on human behavior is essential for creating spaces that are not only functional but also pleasant. In other words, anyone who visits should be able to move around and inhabit the space without experiencing stress or undue effort. Therefore, as shown by the literature on crowding perception in Section 2.2, it is important for us to design environments that account for all observable phenomena, such as motion, as well as unseen factors like psychological, cultural, and other individual or environmental aspects. This means we need to consider not only the physical arrangement but also the social and behavioral dimensions.

Current design standards and guidelines heavily depend on specific, predetermined values for factors, such as those related to space, capacity, or design [70], and might not fully align with real-world scenarios. When considering crowding perception, differences in individual physical characteristics, and various social and behavioral aspects, the practical application of these guidelines can be quite different, as the complexity and variability of human experiences might affect the accuracy of straightforward numerical estimates in real-life situations.

Crowd simulation allows designers to meticulously predict and analyze how people will move and interact within a designed space [82]. Through the simulation of various scenarios, architects explore space usage, overcrowding, movement flows, and accessibility. Crowd simulation is a powerful tool for enhancing the user experience of such a location.

However, the approach to conducting crowd simulations can vary. In these simulations, virtual representations of individuals, often referred to as avatars, are employed and can be crafted in diverse manners. One straightforward technique involves utilizing point-like particles, which essentially denote locations in space [31, 58]. While this approach lacks in capturing detailed nuances, the use of circular particle representations for agents has a long-standing history and remains prevalent in the literature, likely making it the most widely adopted method in crowd simulation studies [29, 24].

Another alternative is to use dynamic two-dimensional (2D) avatar representations, which is advantageous for capturing avatar animations and facilitating a diverse variety of characters, but there is still a lack of realism in this option [57, 70, 71].

While most crowd simulations focus on flat, two-dimensional spaces, a new approach is emerging for more realistic avatars with the use of three-dimensional (3D), virtual reality (VR), and augmented reality (AR) to explore the perception of crowding [71]. The 3D approach includes the use of complex shapes to represent agents and plan their movements in a more realistic and three-dimensional way. As noted in the social psychology literature, human behavior can be influenced by the presence of others, even if those others are virtual avatars [84]. Because of this, incorporating realistic avatars into the architectural design process can be beneficial and should be considered standard practice. In the next section, I explore more related work in more realistic contexts.

2.4 Related Work

In the context of virtual environments and built space design, it is essential to explore how different visualization methods and human representations influence perception and interaction. Although extensive research exists in Human-Computer Interaction (HCI) regarding virtual environments, perception, and scale, there is a gap in studies focusing on the design process for built environments. There is growing interest in enhancing digital experiences and their emotional impacts, but fewer studies address the improvement of current building design processes and the role of more realistic approaches as 3D or VR in futuristic design processes [87].

One relevant study is by Obeidat et al. [62], which investigated how first-year design students' opinions of spatial qualities change when various human avatars are included in a specific environment. This study emphasizes how different representations affect users' emotional and psychological responses, focusing on architectural education and how students perceive digitally designed spaces. The authors found that human figures in architectural visualizations improve distance and scale perception, making spatial judgments more accurate. This insight suggests that including realistic avatars in virtual environments could enhance users' perception of crowding and their emotional responses, which aligns with my work examining different avatar types (particle, 2D, 3D) and their impact on crowding perception among a diverse group of participants.

Moreover, Zibrek et al. [88] examined how photorealistic and stylized virtual characters influence social presence and emotional reactions in VR environments. Their findings indicate that photorealism increases the illusion of being in a real space and affects emotional responses based on visual realism. This suggests that the realism of avatars can significantly influence users' perceptions and emotional reactions in virtual environments, which is pertinent to my research on avatar density and representation. However, they also found that the rendering style did not affect the illusion of presence, highlighting the complexity of these interactions.

Furthermore, recent studies have shown that the existence and realism of virtual avatars influence how individuals perceive and act in a setting [84, 32]. For example, Sanz-Blas et al. [67] explored crowding perceptions in leisure travel destinations, focusing on tourists' experiences and their satisfaction, future behavioral intentions, and electronic word-of-mouth behaviors. Their findings reveal that crowding negatively impacts tourist satisfaction and future behavioral intentions, which extends the understanding of crowding perceptions to virtual environments. This is relevant to my study, which examines how avatar density and representation influence user comfort and emotional states, indirectly affecting user satisfaction and engagement in virtual settings.

Additionally, Koilias et al. [40] explored how different parameters of a virtual crowd—such as density, speed, and direction—affect human movement behaviors within a virtual environment. They found that virtual crowd density and speed significantly impact participants' walking speeds, deviations, and trajectory lengths. This highlights the importance of crowd parameters in shaping user experiences in virtual environments, aligning with my research focus on how individuals interact with virtual crowds.

In a related vein, Berton et al. [9] studied the impact of crowd density on where people look and how they move, using VR to examine human locomotion in dynamic environments. Their research showed that as crowd density increases, participants' focus remains frequent but narrows to concentrate more on individuals directly in front. This finding is relevant as it underscores how crowd density affects visual attention and spatial navigation.

Furthermore, Schwartz et al. [71] examined the effects of different visualization methods on space planning in architecture, focusing on how human form representations impact designers' and the general public's perception of space occupancy. They found that visualization type significantly affects occupancy estimates, with integrated manikins preferred for estimating space usage. This study provides foundational insights for developing design tools that better integrate human factors, paralleling my work on avatar density and representation in virtual environments and their impact on user experience. While their research focused on physical space planning, my study extends this by examining emotional responses and crowding perceptions in virtual spaces, thereby broadening the understanding of human-environment interactions.

In summary, the existing literature demonstrates the critical role of visualization and human representation in shaping perceptions and behaviors in both physical and virtual environments. These studies collectively inform and support the objectives of my research, which aims to deepen our understanding of how avatar density and representation influence crowding perception and emotional responses in virtual environments. By building on these insights, my work contributes to a more comprehensive approach to designing and evaluating virtual spaces, enhancing user comfort, satisfaction, and engagement.

Chapter 3

Methodology

In this chapter, I present the user study conducted as part of my master's research. In Section 3.1, I first discuss the research gaps and objectives, highlighting their relevance in relation to existing literature. Then, in Section 3.2, I provide an overview of the methodology used in this study, including experiment design, data collection, and analysis techniques employed. Finally, I briefly touch on ethical approval and compliance procedures in Section 3.8.

3.1 Study Overview

The main objective of my study is to investigate how different types of avatar representations (particle, 2D, or 3D) and various densities of avatars in a virtual environment affect users' perception of crowding and their emotional responses. Specifically, I aim to examine if people's perceptions and emotional responses differ with crowds of different densities represented by various types of avatars, identify thresholds where significant changes occur, and discover the overall relationship between the number density of avatars and crowding perception by participants. I conducted a preliminary investigation to gain more insights into crowding and its influence on emotions. The following research questions and theories serve as a guide for this thesis:

RQ1: How consistent are users' crowding perceptions and emotional responses (valence, arousal, dominance) when exposed to varying crowd densities and different avatar representations (particle, 2D, and 3D) multiple times?

$RQ1.H_0$: If users are exposed to varying crowd densities with different avatar representations (particle, 2D, and 3D), their crowding perceptions and emotional responses (valence, arousal, and dominance) will remain consistent across multiple exposures.

$RQ1.H_1$**:** If users are exposed to varying crowd densities with particle avatar representations, their crowding perceptions and emotional responses (valence, arousal, and dominance) will significantly differ across multiple exposures.

$RQ1.H_2$**:** If users are exposed to varying crowd densities with 2D avatar representations, their crowding perceptions and emotional responses (valence, arousal, and dominance) will significantly differ across multiple exposures.

$RQ1.H_3$**:** If users are exposed to varying crowd densities with 3D avatar representations, their crowding perceptions and emotional responses (valence, arousal, and dominance) will significantly differ across multiple exposures.

RQ2: Is there a threshold density at which users' emotional responses significantly change?

$RQ2.H_0$**:** If avatars are represented in different formats, the thresholds for significant changes in users' emotional responses will remain the same regardless of the representation (particle, 2D, or 3D).

$RQ2.H_1$**:** If particle representations of avatars are used, the threshold density for significant changes in users' emotional responses will be higher compared to 2D avatar representations.

$RQ2.H_2$**:** If particle representations of avatars are used, the threshold density for significant changes in users' emotional responses will be higher compared to 3D avatar representations.

$RQ2.H_3$**:** If 2D avatars are used, the threshold density for significant changes in users' emotional responses will be higher compared to 3D avatar representations.

RQ3 Do different types of avatar representations (particle, 2D, and 3D) and varying densities of avatars in a virtual environment lead to differences in users' perceptions of crowding and their emotional responses (valence, arousal, and dominance)?

$RQ3.H_0$**:** If different types of avatar representations (particle, 2D, and 3D) are placed in an environment, there will be no significant differences in users' perceptions of crowding and their emotional responses (valence, arousal, and dominance).

$RQ3.H_1$: If particle representations of avatars are placed in an environment, users' perceptions of crowding and their emotional responses (valence, arousal, and dominance) will be less intense compared to when 2D avatars are used.

$RQ3.H_2$: If particle representations of avatars are placed in an environment, users' perceptions of crowding and their emotional responses (valence, arousal, and dominance) will be less intense compared to when 3D avatars are used.

$RQ3.H_3$: If 2D avatars are placed in an environment, users' perceptions of crowding and their emotional responses (valence, arousal, and dominance) will be less intense compared to when 3D avatars are used.

RQ4: Is there a linear relationship between avatar density and perceived crowding, and how do emotions (valence, arousal, dominance) vary linearly with avatar density for different types of avatar representations?

$RQ4.H_0$: If avatar density increases, there will be no linear relationship between avatar density and perceived crowding, nor will there be a linear variation in emotional responses (valence, arousal, and dominance) for different types of avatar representations (particle, 2D, and 3D).

$RQ4.H_1$: If avatar density increases, there will be a linear relationship where perceived crowding increases for particle representations of avatars.

$RQ4.H_2$: If avatar density increases, there will be a linear relationship where perceived crowding increases for 2D representations of avatars.

$RQ4.H_3$: If avatar density increases, there will be a linear relationship where perceived crowding increases for 3D representations of avatars.

$RQ4.H_4$: If avatar density increases, the emotional responses (valence, arousal, and dominance) will vary linearly for particle representations of avatars.

$RQ4.H_5$: If avatar density increases, the emotional responses (valence, arousal, and dominance) will vary linearly for 2D representations of avatars.

$RQ4.H_6$: If avatar density increases, the emotional responses (valence, arousal, and dominance) will vary linearly for 3D representations of avatars.

By conducting this research, I aim to understand how virtual environments and avatar representations influence user experiences. This research will help design more effective and user-friendly virtual spaces, develop guidelines to enhance user comfort and emotional well-being, and improve architectural design processes. Additionally, the study will investigate how incorporating avatars can enhance collaborative design by better addressing user needs and perceptions.

3.2 Study Design

Our research used a within-subjects design, in which all participants underwent the same conditions of experimentation that varied by representation and density of avatars. The general flow of the experiments was as follows, see also Figure 3.1.

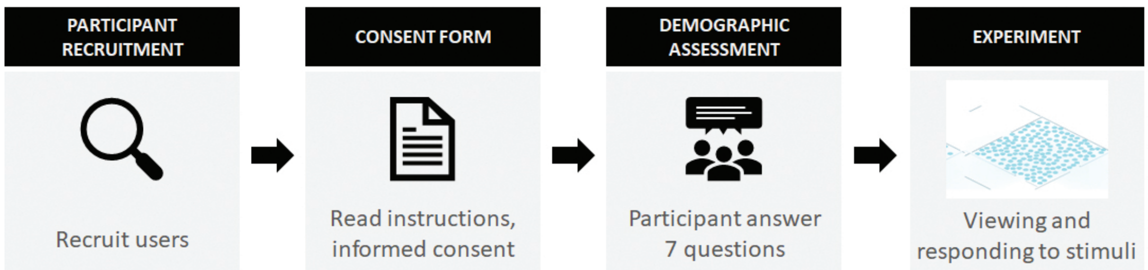

Figure 3.1: Flow of user study. The first step, titled "Participant Recruitment," involves recruiting users. The second step, "Consent Form," is the process where the user reads the form and gives consent to proceed with the experiment. The third step, "Demographic Assessment," includes seven demographic questions that the participant needs to answer before proceeding to the next step. The final step, "Experiment," is the main part of the study, where the user views 90 images and responds to 360 questions.

To gather participants, we utilize multiple recruitment strategies. We sent out emails through a University of Victoria (UVic) email list. Additionally, snowball recruitment was employed to broaden the participant pool; this involved asking recipients of the email invitation to forward it to others who might also want to participate. We also used social media to reach more people and get participants from different backgrounds to join.

Upon agreeing to participate, participants had to read a consent form that has been reviewed and approved by the University of Victoria Ethics Board. This form describes the study and guarantee that participants understand everything before they begin. To proceed, participants must agree to the study and confirm that they have the age of 18 or older. If they qualify, they will begin the main survey on Qualtrics.

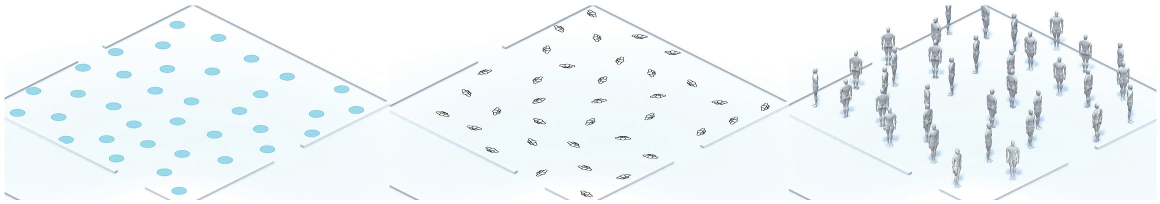

Figure 3.2: The image shows three types of stimuli representing the same environment. From left to right: (a) a 2-dimensional room seen from above populated with particle avatars; (b) a 2-dimensional room seen from above using 2D avatars; and (c) a 2-dimensional room from above using 3D avatars.

All participants accessed the survey through a desktop view, as Qualtrics did not allow individuals using mobile devices to participate in the study. The survey consists of viewing stimuli depicting a room filled with avatars. There are three types of stimuli representing the same environment, as shown in Figure 3.2: (a) 2-dimensional rooms seen from above populated with particle avatars; (b) 2-dimensional rooms seen from above but using 2D avatars instead; and finally (c) the third condition shifts to the same 2-dimensional room from an aerial perspective, filled with 3D avatars.

3.2.1 Densities

The study looks at ten different crowd sizes: 3, 5, 8, 13, 22, 36, 60, 98, 162, and 266 people. These sizes show how crowded a place can be, from very empty to very full. For each crowd size, the study uses three types of avatars: particles, 2D avatars, and 3D avatars. Table 3.2 illustrates the different conditions tested.

Each of the 30 unique combinations (from 10 crowd sizes and 3 avatar types) is shown to participants three times. Repeating them helps make sure the results are accurate and not affected by random answers. So, participants see a total of 90 different images (30 combinations x 3 times each).

To make sure the results are accurate, with good reliability, the order and density in the images are shown in random order. This randomness is important because it stops participants from guessing patterns or getting used to the pictures. By changing the sequence and crowd sizes randomly, the study gets honest reactions about crowding, making the findings more useful for real-life situations.

This approach to examining different densities and avatar representations allows the study to investigate how various factors contribute to the perception of crowding. The diverse range of densities, combined with the detailed and varied avatar representations, creates a detailed set of data that can give us a better understanding of how people perceive crowded spaces and what that means for designing buildings and environments.

Density (people)	**Particle**	**2D Avatar**	**3D Avatar**	**Total Stimuli**
3	Yes	Yes	Yes	2
5	Yes	Yes	Yes	2
8	Yes	Yes	Yes	2
13	Yes	Yes	Yes	2
22	Yes	Yes	Yes	2
98	Yes	Yes	Yes	2
266	Yes	Yes	Yes	2
Total (per variation)	**7**	**7**	**7**	**14**

Table 3.1: Breakdown of stimuli by density and avatar representation. Each density is shown in three variations (particle, 2D, and 3D avatars). Since each stimulus is presented two times, the total number of stimuli shown in the study is 42 (14 stimuli x 2 repetitions).

Density (people)	**Particle**	**2D Avatar**	**3D Avatar**	**Total Stimuli**
3	Yes	Yes	Yes	3
5	Yes	Yes	Yes	3
8	Yes	Yes	Yes	3
13	Yes	Yes	Yes	3
22	Yes	Yes	Yes	3
36	Yes	Yes	Yes	3
60	Yes	Yes	Yes	3
98	Yes	Yes	Yes	3
162	Yes	Yes	Yes	3
266	Yes	Yes	Yes	3
Total (per variation)	**10**	**10**	**10**	**30**

Table 3.2: Breakdown of stimuli by density and avatar representation. Each density is shown in three variations (particle, 2D, and 3D avatars). Since each stimulus is presented three times, the total number of stimuli shown in the study is 90 (30 stimuli x 3 repetitions).

3.2.2 Avatar Representations

The study uses three types of avatar representations to see how different looks affect the crowding perception.

Firstly, particles are utilized due to their widespread application in crowd simulation and architectural contexts [30, 34, 81]. These simple representations are frequently employed to illustrate crowd density because they effectively convey the number of individuals in a given space without the complexities associated with detailed human figures. Figure 3.3 shows the stimuli for particles with densities of 3, 36, and 162.

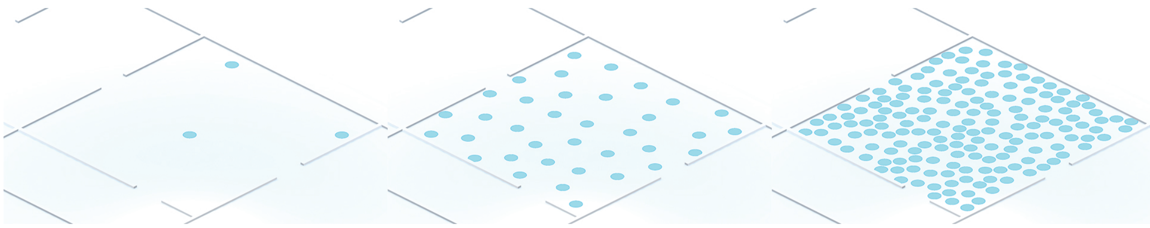

Figure 3.3: The image shows three sections illustrating particle avatars with different crowd densities. The first section on the left has a few particles, representing a density of 3. The middle section has more particles, representing a density of 36. The section on the right is densely packed with particles, representing a density of 162. Each section depicts a 2-dimensional room seen from above, using simple blue dots to represent individuals in the space.

Secondly, the study incorporates 2D avatars, which are a standard reference for human dimensions in architectural design [86]. These avatars offer a more relatable yet still somewhat abstract representation of people. By using 2D avatars, the study bridges the gap between the simplicity of particle representations and the complexity of fully-rendered 3D figures. 2D avatars allow participants to better visualize the presence of other individuals in the space, which can influence their crowding perception. Figure 3.4 shows the stimuli for 2D with densities of 3, 36, and 162.

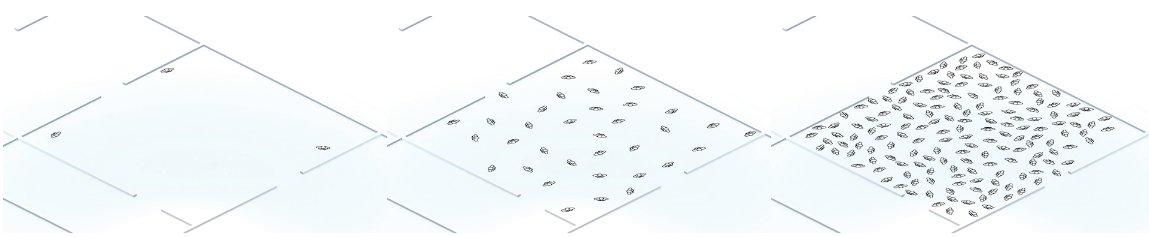

Figure 3.4: The image shows three pictures with different crowd densities of 2D avatars to represent people. The first picture on the left has a few avatars, representing a density of 3. The middle picture has more avatars, representing a density of 36. The picture on the right is full of avatars, representing a density of 3162. Each picture looks down from above and uses simple 2D shapes to show the crowd.

Lastly, the study employs 3D avatars to provide the most realistic portrayal of the human form. These avatars, similar to those used in virtual reality studies, offer detailed and lifelike representations of people [69, 32]. Previous research has shown that 3D digital modeling facilitates the exploration of functional aspects of design and promotes exploratory creativity. By using 3D avatars, the study aims to capture the most accurate and immersive experience of crowding, reflecting real-world conditions where the depth and volume of human figures significantly influence perception. This approach ensures a human-centered design perspective, enhancing ergonomic assessments and improving simulation accuracy [69]. Figure 3.5 shows the stimuli for 3D with densities of 3, 36, and 162.

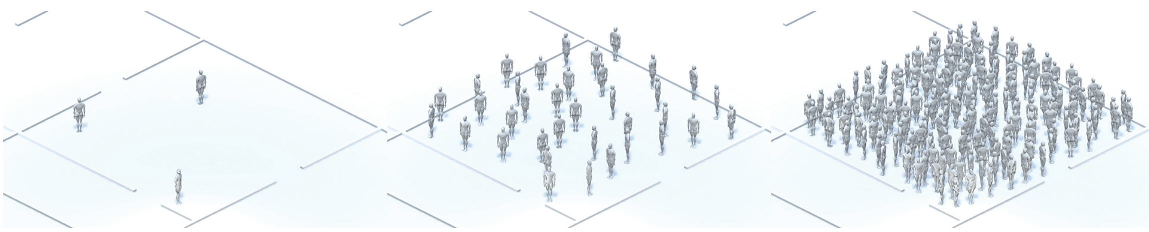

Figure 3.5: The image shows three pictures with different numbers of 3D avatars to represent people. The first picture on the left has a few avatars, showing 3 people. The middle picture has more avatars, showing 36 people. The picture on the right is full of avatars, showing 162 people. Each picture looks down from above and uses detailed 3D shapes to show the crowd.

Each type of avatar representation provides a different level of detail and realism, allowing the study to assess how these variations influence participants' crowding perception. The use of particles, 2D avatars, and 3D avatars in varying densities and contexts enables a comprehensive examination of the factors that contribute to the feeling of being in a crowded environment. This layered approach ensures that the findings are robust and applicable to a range of scenarios, from abstract planning stages to detailed, realistic simulations.

3.2.3 Survey for Measuring Crowding Perception

Participants are asked to report how crowded they perceive each environment to be. For this purpose, the Likert scale was utilized. The Likert scale is a research technique frequently used in perception and attitude studies to measure people's reactions to concepts or objects [27, 61]. Developed by Rensis Likert in the 1930s [48], this scale uses a series of statements or questions with graduated response options to assess respondents' opinions. Participants rated the number of avatars as they viewed each stimulus, ranging from "not crowded at all" to "way too crowded" (1−5), as can be seen in Figure 3.6. This method is widely used to measure perceived crowding [5].

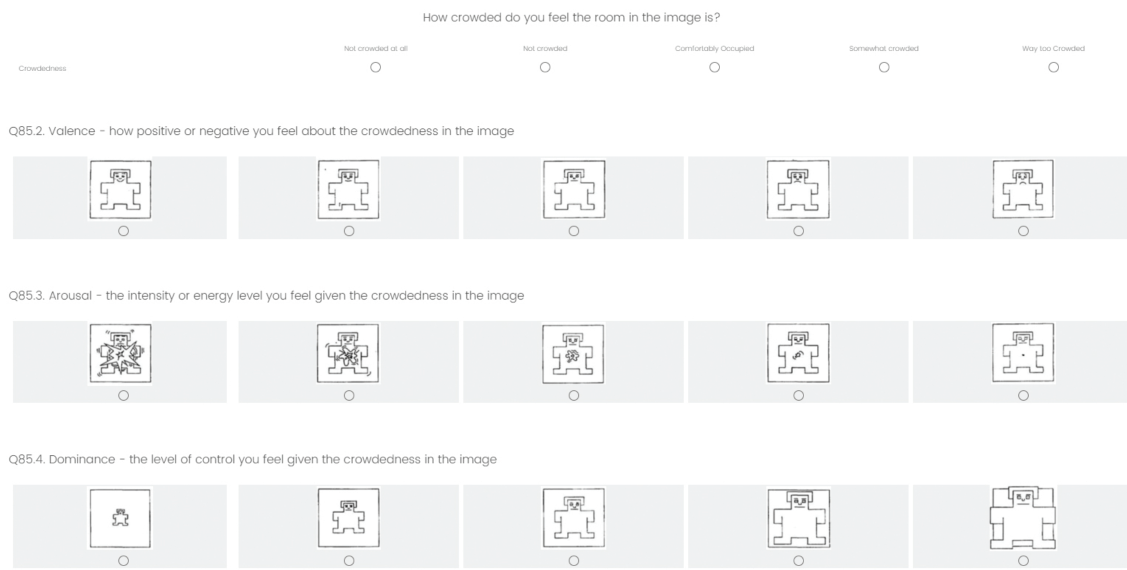

Figure 3.6: The image shows a survey interface with four sections, each containing questions about crowding and emotional responses. A horizontal Likert scale asks participants to rate how crowded they feel the room in the image is. Options range from "Not crowded at all" to "Way too Crowded." A row of five cartoon faces represents the valence scale, asking participants to rate how positive or negative they feel about the crowdedness in the image. The faces range from very unhappy to very happy. Another row of five cartoon faces represents the arousal scale, asking participants to rate the intensity or energy level they feel given the crowdedness in the image. The faces range from very calm to very excited. The last row of five cartoon faces represents the dominance scale, asking participants to rate the level of control they feel given the crowdedness in the image. The faces range from feeling very powerless to very powerful.

The survey also includes subsequent questions designed to assess emotional responses to crowding using three specific dimensions: valence (positive or negative feelings), arousal (intensity of energy level), and dominance (level of control). The Self-Assessment Manikin (SAM) scale, based on Bradley's 1994 study [11], is used to assess emotional responses. Table 3.3 illustrates the different scales from 1 to 5 [46].

The SAM scale uses sequences of humanoid figures to represent gradations along three bipolar affect dimensions:

Scale	Valence	Arousal	Dominance	Crowding
1	Unpleasant	Calm	Controlled	Not crowded at all
2	Unsatisfied	Dull	Powerlessness	Not crowded
3	Neutral	Neutral	Neutral	Comfortably Occupied
4	Pleased	Wide-awake	Powerful	Somewhat crowded
5	Pleasant	Excited	In-control	Way too Crowded

Table 3.3: The table categorizes different scales from 1 to 5, providing descriptions for each scale in terms of four dimensions: Valence, Arousal, Dominance, and Crowding. This table is used to rate different subjective experiences across the dimensions of emotional valence (pleasure or displeasure), arousal (calm or excited), dominance (feeling in control or not), and crowding perception [11, 46].

- **Arousal:** Ranges from feeling active (e.g., alert, excited) to feeling inactive (e.g., uninterested, bored).

- **Valence:** Ranges from unpleasant feelings (e.g., sad, stressed) to pleasant feelings (e.g., happy, elated) .

- **Dominance:** Ranges from feeling helpless and out of control to feeling empowered and in control.

Participants are instructed to select the range for arousal, dominance, and valence that best represents their feelings while viewing an image of a crowded space. SAM scale is easy to administer, non-verbal, and allows for the quick assessment of pleasure, arousal, and dominance in response to an event. Additionally, it is well-validated in the literature [64, 10, 56, 39].

For each stimulus, participants were required to answer four questions. The first question assessed their perception of crowding using a Likert scale. The remaining three questions evaluated their emotional responses in terms of valence, dominance, and arousal. Given that there were 90 stimuli in the study, each participant had to respond to a total of 360 questions.

3.3 Consistency of Perceptions and Responses

In this chapter, I discuss the statistical methods that will be used to analyze the consistency of participants' perceptions of crowding and emotional responses (valence, arousal, dominance) when exposed to different crowd densities and avatar representations (particle, 2D, 3D) in multiple repetitions. Consistency is crucial to ensure the reliability of the collected data and the validity of the study results. Consistency is key to designing reliable and predictable virtual environments. By understanding this, we can ensure that virtual spaces are experienced similarly by all users, every time. Also, when different architects, engineers, and other stakeholders, including the client, who may not be a specialist, are collaborating on a visual representation, it is important that they all perceive it consistently. This consistency helps ensure effective collaboration and accurate feedback during the design process.

To address the consistency of the answers, we will use several complementary statistical methods to analyze participants' responses.

First, we will use the Friedman test, a non-parametric test that detects significant differences in repeated measures within the same group [73]. In this study, the Friedman Test will help us see if there are significant differences in participants' perceptions of crowding and emotional responses between different repetitions of the same stimuli. The test assumes there are no significant differences between the conditions being compared. If we reject this null hypothesis, it indicates that at least two conditions differ significantly. This will help us identify any variations in responses that may indicate inconsistency.

A few examples in the literature about the use of the Friedman Test include a study on soccer players that used the Friedman Test to evaluate changes in imagery factors before and after training with self-modeling videos [3]. This demonstrates the test's broader applicability in analyzing repeated measures of performance-related visual data. In another study about Covid, the test was applied to analyze repeated measures of imaging data across multiple time points [63].

3.4 Creating composite scores from multiple variables

In our experiments, each user analyzed the same stimuli three times. This repetition provided us with a robust dataset and a significantly larger number of responses for each stimulus. To make the best use of this extensive data, it is essential to apply the correct methodology. This is where composite scores come in.

A composite performance measure is a combination of two or more component measures into a single performance measure [59, 35, 4, 75]. The idea of composite measures is to combine multiple measures using a predetermined weighting methodology to produce a single score or variable, facilitating comparison. In the context of this work, this means combining the variables of users rating the same image and stimuli into a single one if it passes the methodology of good reliability. Moreover, the use of composite scores is crucial for answering our specific research questions:

For "RQ2: Is there a threshold density at which users' emotional responses significantly change¿', combining responses into composite scores allows us to more precisely identify if and when significant changes in emotional responses occur at various density levels.

For "RQ3: Do different types of avatar representations (particle, 2D, and 3D) and varying densities of avatars in a virtual environment lead to differences in users' perceptions of crowding and their emotional responses (valence, arousal, and dominance)¿', combining responses into composite scores allows us to effectively compare the impact of different avatar types and densities on user perceptions and emotional responses.

For "RQ4: Is there a linear relationship between avatar density and perceived crowding, and how do emotions (valence, arousal, dominance) vary linearly with avatar density for different types of avatar representations¿', combining responses into composite scores helps us to analyze and understand the potential linear relationships between these variables more accurately.

To ensure the reliability and validity of the composite scores, we used the Intraclass Correlation Coefficient (ICC) (C,k), which measures the consistency of ratings made by the same rater across multiple instances [41]. Interrater reliability refers to how consistent different individuals are at measuring the same phenomenon [7, 83]. To measure the consistency and agreement of participants' responses over repetitions, we calculated the ICC for each combination of density and avatar type, considering crowding perception responses and the three emotional dimensions (valence, arousal, dominance) [13]. High ICC values indicate high consistency in responses over repetitions. Table 3.4 provides a guideline for interpreting ICC values.

ICC Value Range	**Interpretation**
Less than 0.5	Poor reliability
0.5 to 0.75	Moderate reliability
0.75 to 0.9	Good reliability
Greater than 0.9	Excellent reliability

Table 3.4: Interpretation of Intraclass Correlation Coefficient (ICC) Values [41]

The use of ICC ensures the data elements are repeatable and precise. Intrarater variability checks the consistency of assessments made by the same rater or observer over time, measuring how much the same rater agrees with themselves when they assess the same phenomenon at different times [45]. This is crucial for ensuring the reliability of our composite scores. On the other hand, interrater variability refers to the degree of agreement or consistency between different raters or observers when they assess the same phenomenon independently [14, 25]. It measures how much different raters agree on their assessments, which was tested for RQ1 using the Intraclass Correlation Coefficient (ICC) and Friedman Test.

3.5 Threshold Density

To evaluate thresholds and identify significant changes in data, I utilized the change-point detection techniques Binary Segmentation (Binseg) and Pruned Exact Linear Time (PELT). Both methods yielded identical results; thus, I will report the findings from Binseg for simplicity.

Change points signify significant changes in statistical features like average, dispersion, or distribution. Detecting these points helps identify when data behavior or trends shift significantly. In this study, a detected change point indicates a substantial shift in participants' reactions to stimuli at that density, affecting their arousal, valence, crowding, or dominance.

Crowding occurs when density levels impact individuals' behavior, physiology, or experience, influenced by personal traits and social connections [54]. Understanding this transition and its impact on emotional responses is crucial for designing effective virtual environments. Our research question (RQ2) is: Is there a threshold density at which users' emotional responses significantly change? In this section, I outline the methods used to calculate the threshold density at which significant changes in emotional responses (valence, arousal, dominance) occur when participants are exposed to different crowd densities and avatar representations (particle, 2D, 3D). Identifying these thresholds is essential for understanding how varying crowd levels impact emotional responses and informing design decisions in virtual environments.

3.5.1 Methods in a Glance

- Data Preparation: Organize the collected data into a suitable format, including participants' emotional responses (valence, arousal, dominance) at various crowd densities for each type of avatar representation.
- Descriptive Statistics: Calculate the mean, median, and standard deviation for the emotional responses at each crowd density level to provide an initial overview of the data distribution and variability.
- Identifying Potential Thresholds: Examine the descriptive statistics to spot any abrupt changes or trends across the different crowd densities, identifying potential thresholds where significant changes in emotional responses occur.
- Change Point Analysis: Segment the data based on different crowd densities and apply a change point detection algorithm such as the Pruned Exact Linear Time (PELT) or Binary Segmentation to identify the densities where significant changes in emotional responses occur.

In summary, this method details a structured and methodical process to uncover the threshold densities for emotional response changes.

3.6 Influence of Avatar Representations and Densities

Understanding how different avatar representations (particle, 2D, and 3D) and varying avatar densities affect users' perceptions of crowding and their emotional responses (valence, arousal, dominance) requires a detailed and methodical approach. To uncover these insights, I utilized the Friedman Test and the Wilcoxon Signed-Rank Test to reveal the underlying patterns and relationships [73].

The Friedman Test, a non-parametric method, is ideal for analyzing our data as it allows for the comparison of medians across multiple related groups. This test helps identify significant differences in perceptions of crowding and emotional responses across different densities and avatar representations.

When the Friedman Test indicates significant differences, further investigation is necessary to pinpoint exactly which groups differ. This is where the Wilcoxon Signed-Rank Test comes in. This test allows for pairwise comparisons, helping to identify specific pairs of conditions that exhibit significant differences in their medians.

3.6.1 Methods in a Glance

- Friedman Test: Compare the medians of perceptions of crowding and emotional responses (valence, arousal, dominance) across different densities and avatar representations to determine if there are statistically significant differences.

- Wilcoxon Signed-Rank Test: Conduct post-hoc analysis following the Friedman Test to identify which specific pairs of groups differ significantly in their medians. This helps pinpoint the exact conditions that show significant differences.

3.7 Relationship Between Density, Perceived Crowding, and Emotions

To understand the relationship between crowd density, perceived crowding, and emotional responses (valence, arousal, and dominance), I conducted a correlation analysis. This method measures the strength and direction of the relationships between these variables, providing insights into how changes in one variable may be associated with changes in another.

Pearson Correlation Coefficient was used to measure the strength and direction of the linear relationship between responses. This analysis provided additional insights into the relationships between crowd density, perceived crowding, and emotional responses [68].

3.8 Ethical Approval and Compliance

The research protocol for this study, titled "Perception of Environments and Virtual Avatars" (Protocol number: 21-0203), was submitted to the Human Research Ethics Board at the University of Victoria for review. The submission included detailed descriptions of the study's objectives, methodology, participant recruitment strategies, and measures to ensure participant confidentiality and informed consent. After a thorough evaluation, the Ethics Board approved the study, confirming that it met all ethical standards and guidelines for conducting research involving human participants.

In compliance with ethical guidelines, the study participants willingly gave their written consent to join the research study.

Chapter 4

Results Analysis

In this chapter, I share my research findings. I explore how different types of avatars and their densities affect users' emotions and perceptions of overcrowding. By looking at these factors, I discovered new insights and broader implications.

4.1 Study Execution & Participants

This study employed a mixed-methods approach, consisting of two main phases: a pilot study and a main study. The pilot study aimed to test the feasibility and validity of the experimental design and measures used, while the main study examined the effects of different avatar representations and densities on the perception of crowding in a virtual environment.

Participants were recruited through social media and an email list from the University of Victoria (Uvic). Additionally, we used the snowball recruitment method, asking recipients to forward the invitation to others who might be interested in participating. All participants were over 18 years old and provided informed consent before participating.

The study was conducted via an online survey hosted on the Qualtrics platform. A total of 70 participants took part in the pilot study, while 48 participated in the main study. Using a within-subjects design, all participants were exposed to the same experimental conditions, which varied in avatar representation and density. Participants saw three types of avatars: particle avatars (simplified crowd simulations), 2D avatars (standard architectural representations), and 3D avatars (three-dimensional human forms).

In the main study, avatar densities ranged from a maximum of 266 people to a minimum of 5 people, covering ten different intensities for each of the three avatar types. This resulted in 30 unique images. Each image was associated with four questions, assessing crowding, valence, arousal, and dominance, leading to 120 evaluations in total. To ensure consistency in participants' responses, each image was presented three times, resulting in 360 evaluations. The order and density of the stimuli were randomized to avoid learning effects between conditions.

Participants viewed a single virtual environment, specifically a room from the MET, filled with avatars according to the experimental conditions. After viewing, they reported their perception of crowding using a semantic differential scale, similar to a Likert scale but with polarized attributes. Additionally, we measured participants' emotional responses using the Self-Assessment Manikin (SAM) scale, which assesses three emotional dimensions: valence (pleasant to unpleasant), arousal (excited to calm), and dominance (controlled to in control).

By combining crowding analysis with emotional responses using the Self-Assessment Manikin, our study aimed to provide a comprehensive understanding of how different avatar representations and densities affect participants' perception of crowding and emotional responses. This robust methodological approach ensured the validity and reliability of the findings, significantly contributing to the existing literature on the perception of virtual environments and architectural design.

4.1.1 American Psychological Association Style

To increase the ease of reading and comprehension of the experiments results, I am following the American Psychological Association Style is this work. The guidelines originated in 1929 and have been evolving in a iterative effort of the research community since them, is a set of guidelines for writing research papers and scholarly documents. It is widely used in the social sciences, education, engineering, and business fields [8, 1].

4.1.2 Pilot Study

The pilot study ran from March 28, 2023, to October 12, 2023. Conducted online using the Qualtrics platform, the study included 70 participants. Using a within-subjects design, participants were exposed to experimental conditions varying in avatar representation (particle, 2D, and 3D) and density. The valid densities evaluated in the pilot study were 3, 5, 8, 13, 22, 162, and 266, resulting in 27 different stimuli and images. Unlike the main study, each stimulus in the pilot was shown to participants twice, not three times.

Each image was evaluated for crowding, valence, arousal, and dominance to ensure response consistency. The virtual environment viewed was a room filled with avatars. Perception of crowding was assessed using a semantic differential scale, and emotional responses were measured with the Self-Assessment Manikin (SAM) scale. With these settings, the pilot study was successfully deployed and the findings will be showed below.

Participants

A total of 70 participants completed the survey. The sample consisted of 56 females (80%) and 14 males (20%).

Assigned Sex at Birth	n	%
Female	56	80
Male	14	20

Table 4.1: Assigned Sex at Birth of Participants (N = 70)

Education Level

Participants reported various levels of education. The majority of participants held a 4-year degree (28.6%) or a professional degree (28.6%). Additionally, 22.9% of participants were currently working on a degree, 11.4% held a 2-year degree, 4.3% had completed a doctorate, and 4.3% were high school graduates.

Education Level	n	%
2-year degree	8	11.4
4-year degree	20	28.6
Degree in progress (currently working on a degree)	16	22.9
Doctorate	3	4.3
High school graduate	3	4.3
Professional degree	20	28.6

Table 4.2: Highest Completed Education Level of Participants (N = 70)

Design Experience

When asked about their design experience, 52 participants (74.3%) reported that they do not consider themselves designers, while 18 participants (25.7%) identified as designers.

Do you consider yourself a designer?	n	%
No	52	74.3
Yes	18	25.7

Table 4.3: Design Experience of Participants (N = 70)

Consistency Analysis for Repeated Stimuli

As mentioned in Chapter 3, non-parametric tests were selected to assess significant differences between groups. To assess significant differences between pairs of groups, the Wilcoxon signed-rank test is applied and usually serves as a post-hoc test for the Friedman Test, the latter being used for comparisons between more than two groups [89].

For the pilot study, the participants only saw the same stimuli twice. For this reason, only the Wilcoxon Test was used to determine if there are differences between the ratings for the same stimuli among the 70 participants or if the ratings were consistent.

Table 4.4 shows the Wilcoxon results for the consistency of responses to the same image among the participants for crowding perception. The Wilcoxon signed-rank test indicated that there was no significant difference between the repeated stimuli for crowding perception across all densities for avatar representations in particles and 3D, with all p-values $> .05$.

For 2D avatar representations, there are statistically significant differences identified by the Wilcoxon signed-rank test. The Wilcoxon signed-rank test results showed that there was a significant difference in density 5 between the same image, $p < .01$, $Z = 69$. The Wilcoxon signed-rank test results for 2D avatar representations at density 8 also indicated a significant p-value, with a significant difference in density 8 between the same image, $p = .03$, $Z = 235.5$.

This suggests that participants who saw the same stimuli twice showed significantly inconsistent ratings for crowding perception at these densities when viewing

Statistic	p_value	Stimuli_type	Density
9	0.15	2d	3
10.5	1.00	3d	3
5	1.00	particle	3
69	0.01**	2d	5
184.5	0.64	3d	5
89	0.33	particle	5
235.5	0.03**	2d	8
189	0.33	3d	8
170	0.40	particle	8
162	0.98	2d	13
110	0.82	3d	13
190	0.34	particle	13
342	0.45	2d	22
170	0.61	3d	22
207	0.57	particle	22
8	0.13	2d	162
5	0.23	3d	162
30	0.76	particle	162
8	0.25	2d	266
0	0.15	3d	266
0	1.00	particle	266

Table 4.4: Results of the Wilcoxon Signed-Rank Test for Crowding Perception. The table presents the results of the Wilcoxon signed-rank test for the consistency of responses for crowding perception at different densities. The table includes the test statistic (Z) and the p-value for each density level. Specifically, the table highlights significant differences (**) in crowding perception for densities 5 and 8 when participants viewed the same stimuli twice.

2D avatar representations.

Table 4.5 shows the Wilcoxon results for the consistency of responses to the same image among the participants for arousal. The Wilcoxon signed-rank test indicated that there was no significant difference between the repeated stimuli for arousal across all densities for avatar representations in particles and 3D, with all p-values $> .05$.

For 2D avatar representations, there are statistically significant differences identified by the Wilcoxon signed-rank test. The Wilcoxon signed-rank test results showed that there was a significant difference in density 5 between the same image, $p = .02$, $Z = 92$. The Wilcoxon signed-rank test results for 2D avatar representations at density 8 also indicated a significant p-value, with a significant difference in density 8 between the same image, $p = .02$, $Z = 71$.

Statistic	p_value	Stimuli_type	Density
74.5	0.62	2d	3
63	0.78	3d	3
13.5	0.07	particle	3
92	0.02**	2d	5
103	0.93	3d	5
54.5	0.05	particle	5
71	0.02**	2d	8
202.5	0.98	3d	8
189	0.33	particle	8
168	0.26	2d	13
172	0.44	3d	13
209	0.42	particle	13
348.5	0.96	2d	22
188	0.49	3d	22
210	0.43	particle	22
96	0.28	2d	162
51	0.37	3d	162
129.5	0.53	particle	162
56.5	0.83	2d	266
19	0.37	3d	266
21	0.14	particle	266

Table 4.5: Results of the Wilcoxon Signed-Rank Test for Arousal. The table presents the results of the Wilcoxon signed-rank test for the consistency of responses for arousal at different densities. The table includes the test statistic (Z) and the p-value for each density level. Specifically, the table highlights significant differences (**) in crowding perception for densities 5 and 8 when participants viewed the same stimuli twice.

This suggests that participants who saw the same stimuli twice showed significantly inconsistent ratings for arousal at these densities when viewing 2D avatar representations.

Table 4.6 shows the Wilcoxon results for the consistency of responses to the same image among the participants for dominance. For 2D avatar representations, there are statistically significant differences identified by the Wilcoxon signed-rank test for dominance. The Wilcoxon signed-rank test results showed that there was a significant difference in density 22 between the same image, $p = .03$, $Z = 196$. For 3D avatar representations, the Wilcoxon signed-rank test results showed that there was a significant difference in density 22 between the same image, $p = .01$, $Z = 158$. For particle avatar representations, the Wilcoxon signed-rank test results showed that there was

Statistic	p_value	Stimuli_type	Density
166.5	0.58	2d	3
62.5	0.10	3d	3
42	0.04**	particle	3
247	0.74	2d	5
118	0.13	3d	5
118.5	0.14	particle	5
279	0.54	2d	8
255.5	0.12	3d	8
243.5	0.20	particle	8
335	0.79	2d	13
199	0.92	3d	13
239	0.11	particle	13
196	0.03**	2d	22
158	0.01**	3d	22
316.5	0.25	particle	22
75.5	0.14	2d	162
19.5	0.71	3d	162
80	0.52	particle	162
58.5	0.38	2d	266
26	0.87	3d	266
20.5	0.46	particle	266

Table 4.6: Results of the Wilcoxon Signed-Rank Test for Dominance. The table presents the results of the Wilcoxon signed-rank test for the consistency of responses for dominance at different densities. The table includes the test statistic (Z) and the p-value for each density level. Specifically, the table highlights significant differences (**) in crowding perception for densities 3 and 22 when participants viewed the same stimuli twice.

a significant difference in density 3 between the same image, $p = .04$, $Z = 42$.

This suggests that participants who saw the same stimuli twice showed significantly inconsistent ratings for dominance at these densities when viewing 2D, 3D, and particle avatar representations.

Table 4.7 shows the Wilcoxon results for the consistency of responses to the same image among the participants for valence. The Wilcoxon signed-rank test indicated that there was no significant difference between the repeated stimuli for valence across all densities for avatar representations in particles and 3D, with all p-values $> .05$.

For 2D avatar representations, there are statistically significant differences identified by the Wilcoxon signed-rank test. The Wilcoxon signed-rank test results showed that there was a significant difference in density 3 between the same image, $p = .008$,

Statistic	p_value	Stimuli_type	Density
18	0.56	3d	3
18	0.07	particle	3
76	0.008**	2d	3
77	0.27	3d	5
148.5	0.69	particle	5
75	0.02**	2d	5
211	0.27	3d	8
162	0.12	particle	8
293	0.69	2d	8
234	0.36	2d	13
280.5	0.74	3d	13
229	0.32	particle	13
427.5	0.55	2d	22
222	0.16	3d	22
223.5	0.84	particle	22
39	0.62	2d	162
25	0.47	3d	162
91	0.86	particle	162
27	0.32	2d	266
9	0.38	3d	266
4	0.70	particle	266

Table 4.7: Results of the Wilcoxon Signed-Rank Test for Valence. The table presents the results of the Wilcoxon signed-rank test for the consistency of responses for valence at different densities. The table includes the test statistic (Z) and the p-value for each density level. Specifically, the table highlights significant differences (**) in crowding perception for densities 3 and 5 when participants viewed the same stimuli twice.

$Z = 76$. The Wilcoxon signed-rank test results for 2D avatar representations at density 5 also indicated a significant p-value, with a significant difference in density 5 between the same image, $p = .02$, $Z = 75$.

This suggests that participants who saw the same stimuli twice showed significantly inconsistent ratings for valence at these densities when viewing 2D avatar representations.

Creating composite scores from multiple variables

Before combining into a composite variable, guidelines recommend assessing its strength. Following Chapter 3, Intraclass Coefficient Class (ICC) and Cronbach's alpha are used to evaluate the reliability of the data. Both tests yield similar results due to their formula. Therefore, I used the Intraclass Coefficient class (C, k) [80, 80].

Density	Attribute	2d	3d	Particle
3	Crowding Perception	0.804	0.729	0.848
3	Valence	0.763	0.952	0.943
3	Arousal	0.851	0.881	0.943
3	Dominance	0.761	0.879	0.899
5	Crowding Perception	0.764	0.722	0.713
5	Valence	0.827	0.691	0.703
5	Arousal	0.894	0.892	0.873
5	Dominance	0.773	0.838	0.795
8	Crowding Perception	0.713	0.773	0.826
8	Valence	0.754	0.826	0.792
8	Arousal	0.910	0.893	0.867
8	Dominance	0.795	0.818	0.808
13	Crowding Perception	0.837	0.889	0.744
13	Valence	0.781	0.851	0.724
13	Arousal	0.839	0.856	0.786
13	Dominance	0.650	0.830	0.765
22	Crowding Perception	0.787	0.843	0.775
22	Valence	0.647	0.786	0.792
22	Arousal	0.755	0.772	0.824
22	Dominance	0.705	0.731	0.802
162	Crowding Perception	0.771	0.574	0.735
162	Valence	0.813	0.609	0.702
162	Arousal	0.923	0.880	0.864
162	Dominance	0.850	0.955	0.898
266	Crowding Perception	0.658	0.506	0.900
266	Valence	0.770	0.761	0.861
266	Arousal	0.922	0.898	0.890
266	Dominance	0.919	0.930	0.948

Table 4.8: ICC(C, k) for Different Densities and Attributes. The table shows that ICC(C, k) values for different densities and attributes across 2D, 3D, and Particle stimuli types generally indicate high reliability, with most values above 0.7.

Under the assumption of randomly selected raters, the ICC for interrater consistency (C, k) assesses the reliability of the observed differences between subjects by accounting for the variance due to subjects while excluding the rater effects. This means it measures how consistently subjects are rated across different raters, and for average ratings, it divides the error variance by the number of raters to reduce rater-related errors. Essentially, ICC(C, k) for interrater consistency focuses on excluding the variability introduced by different raters rating the same image differently at different times. It aims to measure the consistency of the observed differences between subjects while accounting for the average ratings across multiple raters, thereby reducing the impact of rater-related errors or inconsistencies.

I chose to use ICC for interrater reliability instead of intrarater reliability deliberately because raters' inconsistencies would result in weak reliability. These inconsistencies, however, cannot be excluded from the analysis as they are crucial for addressing the research questions. For the composite score, the focus is not on individual raters' consistency (intrarater) but on the overall consistency across different raters (interrater). This approach ensures that the analysis captures the general reliability of ratings across multiple raters, which is essential for the study.

Koo et al. [42] provide the following guidelines for interpretation: *"Values less than 0.5 indicate poor reliability, values between 0.5 and 0.75 indicate moderate reliability, values between 0.75 and 0.9 indicate good reliability, and values greater than 0.90 indicate excellent reliability."* It is important to assess the reliability of judgments made by participants to determine the extent to which measurements are reliable [74].

The intraclass correlation coefficients (ICCs) for different densities and attributes were calculated across 2D, 3D, and Dot stimuli types. The results in Table 4.8 indicated high reliability across most conditions. Regarding the 2D stimuli type, the ICC values ranged from 0.658 to 0.837 for crowding perception, indicating moderate to high reliability. For valence, the ICC values ranged from 0.647 to 0.827, indicating moderate to high reliability. For arousal, the ICC values ranged from 0.650 to 0.923, indicating moderate to high reliability. For dominance, the ICC values ranged from 0.705 to 0.919, indicating high reliability.

Regarding the 3D stimuli type, the ICC values ranged from 0.506 to 0.889, indicating moderate to high reliability for crowding perception. For valence, the ICC values ranged from 0.609 to 0.952, indicating moderate to high reliability. For arousal, the ICC values ranged from 0.700 to 0.898, indicating moderate to high reliability. For dominance, the ICC values ranged from 0.731 to 0.955, indicating high reliability.

Regarding the particle stimuli type, the ICC values ranged from 0.713 to 0.848, indicating high reliability for crowding perception. For valence, the ICC values ranged from 0.571 to 0.943, indicating moderate to high reliability. For arousal, the ICC values ranged from 0.727 to 0.943, indicating high reliability. For dominance, the ICC values ranged from 0.765 to 0.948, indicating high reliability.

Compared to 2D and 3D stimuli, particle stimuli consistently had higher ICC values, which suggests they were more reliable at measuring crowding perception, valence, arousal, and dominance. Based on these results, a composite variable can be created by combining the two ratings for each stimuli type and density into a single variable.

Descriptive analysis

The descriptive statistics for crowding perception across different densities and stimuli types are summarized in Table 4.9. For the 2D stimuli, the mean crowding perception values ranged from 1.11 ($SD = 0.38$) at a density of 3 to 4.92 ($SD = 0.27$) at a density of 266. For the 3D stimuli, the mean values ranged from 1.09 ($SD = 0.28$) at a density of 3 to 4.99 ($SD = 0.12$) at a density of 266. Lastly, for the particle stimuli, the mean values ranged from 1.09 ($SD = 0.31$) at a density of 3 to 5.00 ($SD = 0.00$) at a density of 266.

	Density						
Stimuli Type	**3**	**5**	**8**	**13**	**22**	**162**	**266**
2D	1.11 (0.38)	1.46 (0.67)	1.96 (0.90)	2.49 (0.86)	3.25 (0.87)	4.81 (0.41)	4.92 (0.27)
3D	1.09 (0.28)	1.54 (0.73)	2.25 (0.86)	2.84 (0.82)	3.41 (0.86)	4.91 (0.38)	4.99 (0.12)
Particle	1.09 (0.31)	1.41 (0.67)	2.01 (0.87)	2.71 (0.81)	3.25 (0.91)	4.82 (0.38)	5.00 (0.00)

Table 4.9: Descriptive statistics for crowding perception across different densities, with mean values and standard deviations in parentheses. The table indicates that mean values generally increase with density for 2D, 3D, and particle stimuli, with standard deviations reflecting varying degrees of variability.

The descriptive statistics for arousal across different densities and stimuli types are summarized in Table 4.14. For the 2D stimuli, the mean arousal values ranged from 2.03 ($SD = 1.46$) at a density of 3 to 3.80 ($SD = 1.60$) at a density of 266. For the 3D stimuli, the mean values ranged from 1.97 ($SD = 1.47$) at a density of 3 to 4.16 ($SD = 1.51$) at a density of 266. Lastly, for the particle stimuli, the mean values ranged from 1.89 ($SD = 1.41$) at a density of 3 to 4.08 ($SD = 1.54$) at a density of 266.

Stimuli Type	Density 3	5	8	13	22	162	266
2D	2.03 (1.46)	2.13 (1.30)	2.23 (1.25)	2.34 (1.13)	2.84 (1.00)	3.74 (1.46)	3.80 (1.60)
3D	1.97 (1.47)	2.01 (1.36)	2.22 (1.14)	2.54 (1.09)	2.90 (1.00)	4.04 (1.51)	4.16 (1.51)
particle	1.89 (1.41)	1.91 (1.31)	2.14 (1.18)	2.56 (1.09)	2.74 (1.10)	3.97 (1.41)	4.08 (1.54)

Table 4.10: Descriptive statistics for arousal across different densities, with mean values and standard deviations in parentheses. The table indicates that mean values generally increase with density for 2D, 3D, and particle stimuli, with standard deviations reflecting varying degrees of variability.

The descriptive statistics for dominance across different densities and stimuli types are summarized in Table 4.15. For the 2D stimuli, the mean dominance values ranged from 3.69 ($SD = 1.42$) at a density of 3 to 2.12 ($SD = 1.56$) at a density of 266. For the 3D stimuli, the mean values ranged from 3.56 ($SD = 1.50$) at a density of 3 to 2.12 (1.69) at a density of 266. Lastly, for the particle stimuli, the mean values ranged from 3.59 ($SD = 1.42$) at a density of 3 to 2.09 ($SD = 1.63$) at a density of 266.

Stimuli Type	Density 3	5	8	13	22	162	266
2D	3.69 (1.42)	3.40 (1.32)	3.23 (1.24)	3.10 (1.11)	2.82 (0.90)	2.22 (1.52)	2.12 (1.56)
3D	3.56 (1.50)	3.33 (1.41)	3.29 (1.15)	3.06 (1.01)	2.71 (0.96)	2.12 (1.66)	2.12 (1.69)
particle	3.59 (1.42)	3.57 (1.34)	3.38 (1.18)	2.93 (1.02)	2.75 (0.98)	2.07 (1.47)	2.09 (1.63)

Table 4.11: Descriptive statistics for dominance across different densities, with mean values and standard deviations in parentheses. The table indicates that mean values generally decrease with density for 2D, 3D, and particle stimuli, with standard deviations reflecting varying degrees of variability.

The descriptive statistics for valence across different densities and stimuli types are summarized in Table 4.16. For the 2D stimuli, the mean valence values ranged from 4.35 ($SD = 1.00$) at a density of 3 to 1.26 ($SD = 0.57$) at a density of 266. For the 3D stimuli, the mean values ranged from 4.46 ($SD = 0.94$) at a density of 3 to 1.12 ($SD = 0.44$) at a density of 266. Lastly, for the particle stimuli, the mean values ranged from 4.46 ($SD = 0.96$) at a density of 3 to 1.12 ($SD = 0.42$) at a density of 266.

Assessing Thresholds

To assess thresholds, the Changepoint Detection method will be used. The Binary Segmentation (Binseg) method allows for precise identification of shifts in ratings,

Stimuli Type	Density						
	3	**5**	**8**	**13**	**22**	**162**	**266**
2D	4.35 (1.00)	4.23 (0.86)	3.95 (0.92)	3.71 (0.88)	3.03 (0.86)	1.39 (0.63)	1.26 (0.57)
3D	4.46 (0.94)	4.26 (0.92)	3.82 (0.92)	3.50 (0.98)	2.93 (1.02)	1.26 (0.70)	1.12 (0.44)
particle	4.46 (0.96)	4.44 (0.74)	3.98 (0.88)	3.56 (0.85)	3.08 (0.95)	1.39 (0.66)	1.12 (0.42)

Table 4.12: Descriptive statistics for valence across different densities, with mean values and standard deviations in parentheses. The table indicates that mean values generally decrease with density for 2D, 3D, and particle stimuli, with standard deviations reflecting varying degrees of variability.

providing insights into how perception varies with different avatar densities in 2D, 3D, and Dot virtual environments. Change Point is a moment of time when a time series changes its behaviour [2].

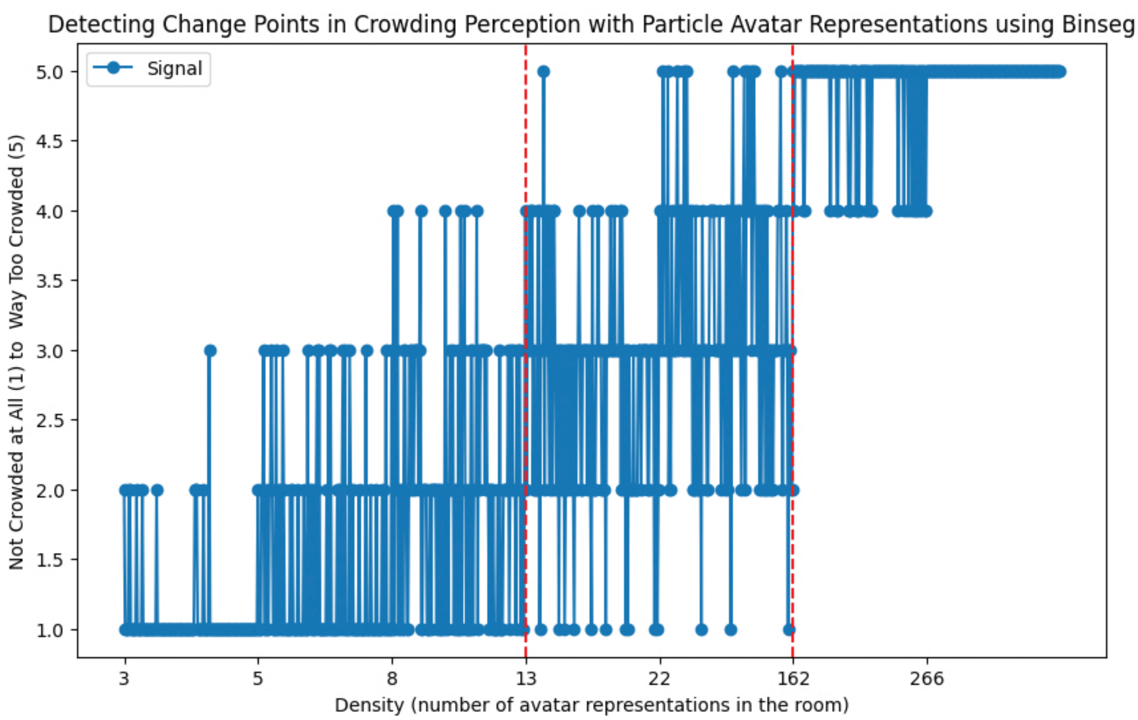

Figure 4.1: Detecting Change Points in Crowding Perception with Particle Avatar Representations using Binseg. The figure illustrates the detection of change points in crowding perception with particle avatar representations using the Binary Segmentation (Binseg) method. The red dashed lines represent the identified change points at densities 13 and 162.

The Figure 4.1 illustrates the detection of change points in crowding perception with particle avatar representations using the Binary Segmentation (Binseg) method.

The x-axis represents the density, while the y-axis represents the crowding perception, ranging from Not Crowded at All to Way Too Crowded (1-5). The blue points indicate the signal, with red dashed lines marking the identified change points at densities 13 and 162.

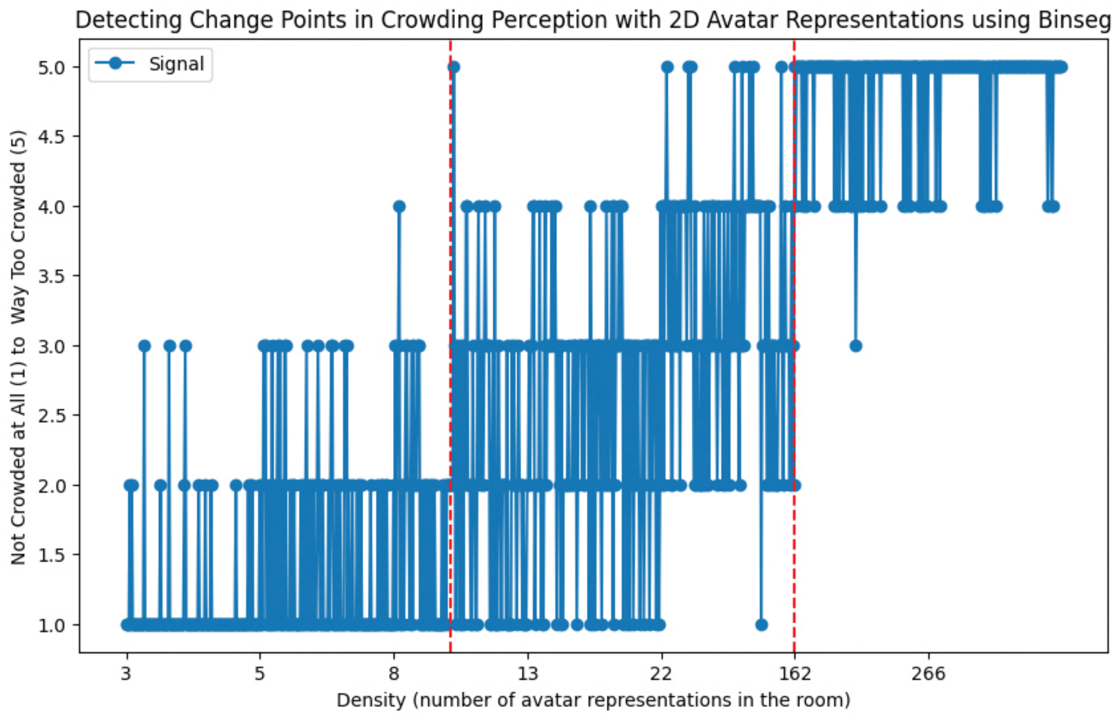

Figure 4.2: Detecting Change Points in Crowding Perception with 2D Avatar Representations using Binseg. The figure illustrates the detection of change points in crowding perception with 2D avatar representations using the Binary Segmentation (Binseg) method. The red dashed lines represent the identified change points at densities 8 and 162.

The Figure 4.2 illustrates the detection of change points in crowding perception with 2D avatar representations using the Binary Segmentation (Binseg) method. The x-axis represents the density, while the y-axis represents the crowding perception, ranging from Not Crowded at All to Way Too Crowded (1-5). The blue points indicate the signal, with red dashed lines marking the identified change points at densities 8 and 162.

The Figure 4.3 illustrates the detection of change points in crowding perception with 3D avatar representations using the Binary Segmentation (Binseg) method. The x-axis represents the density, while the y-axis represents the crowding perception, ranging from Not Crowded at All to Way Too Crowded (1-5). The blue points

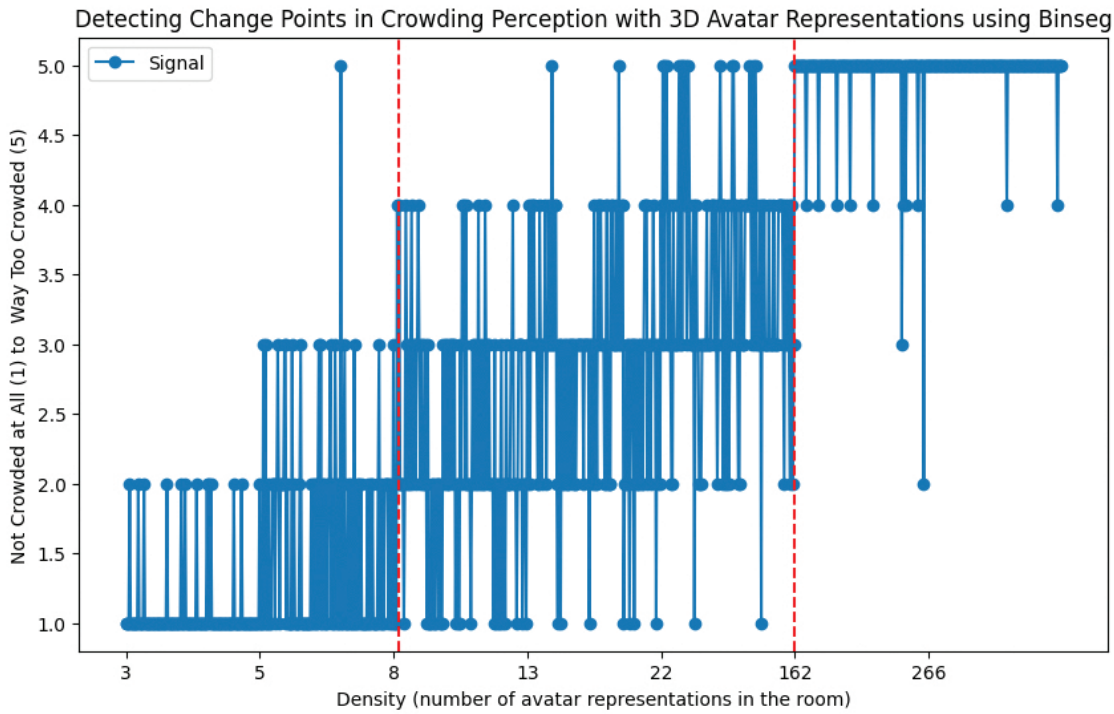

Figure 4.3: Detecting Change Points in Crowding Perception with 3D Avatar Representations using Binseg. The figure illustrates the detection of change points in crowding perception with 3D avatar representations using the Binary Segmentation (Binseg) method. The red dashed lines represent the identified change points at densities 8 and 162.

indicate the signal, with red dashed lines marking the identified change points at densities 8 and 162.

The Figure 4.4 illustrates the detection of change points in arousal with particle avatar representations using the Binary Segmentation (Binseg) method. The x-axis represents the density, while the y-axis represents the arousal, ranging from Calm to Excited (1-5). The blue points indicate the signal, with red dashed lines marking the identified change points at densities 8 and 162.

The Figure 4.5 illustrates the detection of change points in arousal with 2D avatar representations using the Binary Segmentation (Binseg) method. The x-axis represents the density, while the y-axis represents the arousal, ranging from Calm to Excited (1-5). The blue points indicate the signal, with red dashed lines marking the identified change points at densities 8 and 162.

The Figure 4.6 illustrates the detection of change points in arousal with 3D avatar representations using the Binary Segmentation (Binseg) method. The x-axis repre-

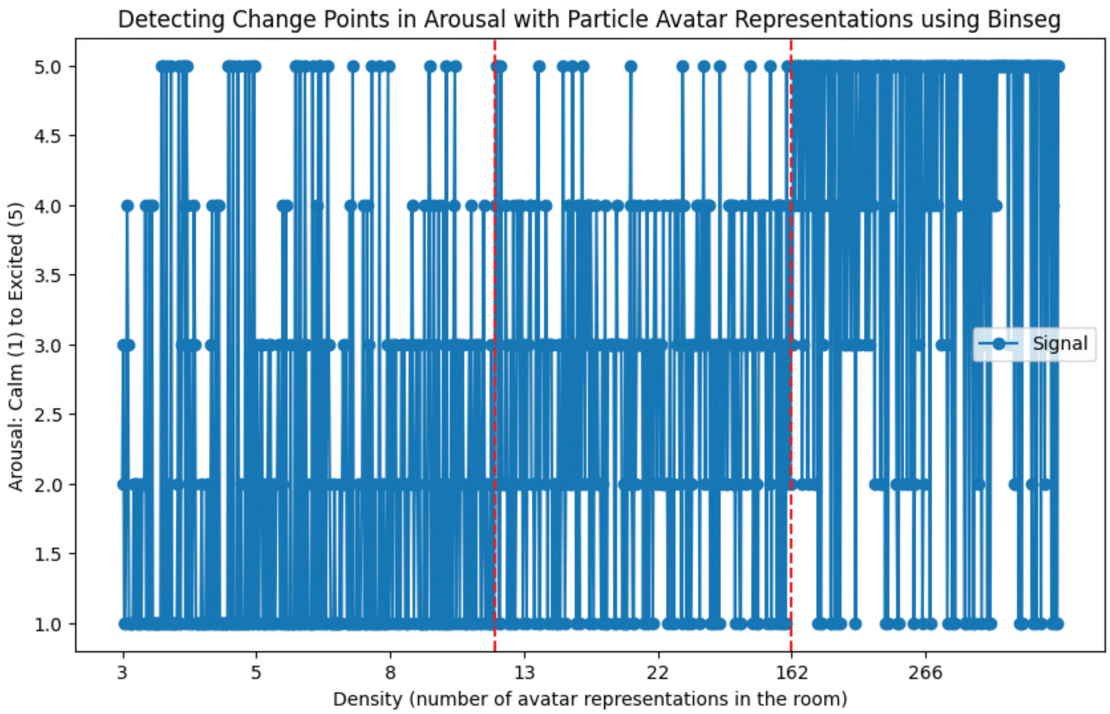

Figure 4.4: Detecting Change Points in Arousal with Particle Avatar Representations using Binseg. The figure illustrates the detection of change points in arousal with particle avatar representations using the Binary Segmentation (Binseg) method. The red dashed lines represent the identified change points at densities 8 and 162.

sents the density, while the y-axis represents the arousal, ranging from calm to excited (1-5). The blue points indicate the signal, with red dashed lines marking the identified change points at densities 8 and 162.

The Figure 4.7 illustrates the detection of change points in dominance with particle avatar representations using the Binary Segmentation (Binseg) method. The x-axis represents the density, while the y-axis represents the dominance, ranging from controlled to In-control (1-5). The blue points indicate the signal, with red dashed lines marking the identified change points at densities 13 and 162.

The Figure 4.8 illustrates the detection of change points in dominance with 2D avatar representations using the Binary Segmentation (Binseg) method. The x-axis represents the density, while the y-axis represents the dominance, ranging from controlled to In-control (1-5). The blue points indicate the signal, with red dashed lines marking the identified change points at densities 8 and 162.

The Figure 4.9 illustrates the detection of change points in dominance with 3D avatar representations using the Binary Segmentation (Binseg) method. The x-axis

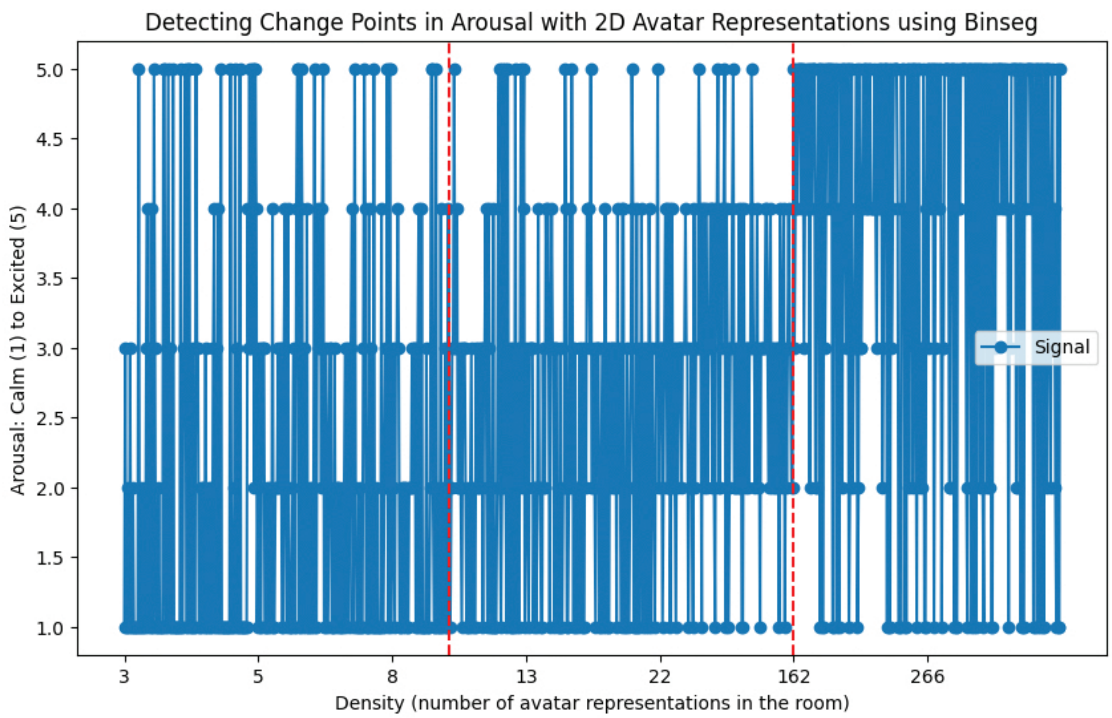

Figure 4.5: Detecting Change Points in Arousal with 2D Avatar Representations using Binseg. The figure illustrates the detection of change points in arousal with 2D avatar representations using the Binary Segmentation (Binseg) method. The red dashed lines represent the identified change points at densities 8 and 162.

represents the density, while the y-axis represents the dominance, ranging from controlled to In-control (1-5). The blue points indicate the signal, with red dashed lines marking the identified change points at densities 8 and 162.

The Figure 4.10 illustrates the detection of change points in valence with particle avatar representations using the Binary Segmentation (Binseg) method. The x-axis represents the density, while the y-axis represents the valence, ranging from unpleasant to pleasant (1-5). The blue points indicate the signal, with red dashed lines marking the identified change points at densities 8 and 162.

The Figure 4.11 illustrates the detection of change points in valence with 2D avatar representations using the Binary Segmentation (Binseg) method. The x-axis represents the density, while the y-axis represents the valence, ranging from unpleasant to pleasant (1-5). The blue points indicate the signal, with red dashed lines marking the identified change points at densities 13 and 162.

The Figure 4.12 illustrates the detection of change points in valence with 3D avatar representations using the Binary Segmentation (Binseg) method. The x-axis repre-

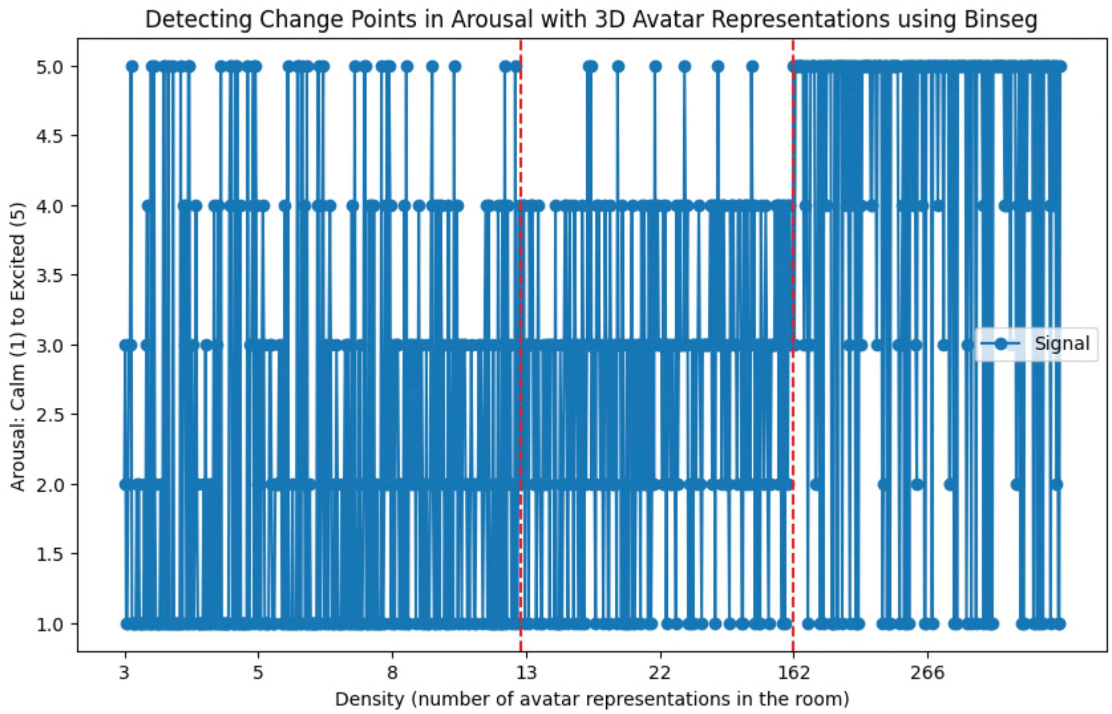

Figure 4.6: Detecting Change Points in Arousal with 3D Avatar Representations using Binseg. The figure illustrates the detection of change points in arousal with 3D avatar representations using the Binary Segmentation (Binseg) method. The red dashed lines represent the identified change points at densities 8 and 162.

sents the density, while the y-axis represents the valence, ranging from unpleasant to pleasant (1-5). The blue points indicate the signal, with red dashed lines marking the identified change points at densities 8 and 162.

Assessing differences between avatar representation

To find out if there are differences between different types of avatar representations, we use the Friedman Test. As mentioned in Chapter 3, the Friedman Test is used to compare different groups to see if they are really different from each other. It's like checking if people have different perceptions for 3D, 2D, and particle avatars. If the Friedman Test tells us that the rankings are different, we then use another test called the Wilcoxon signed-rank test as post-hoc to see exactly which avatars are different from each other. This is done carefully to make sure our findings are accurate and not just by chance.

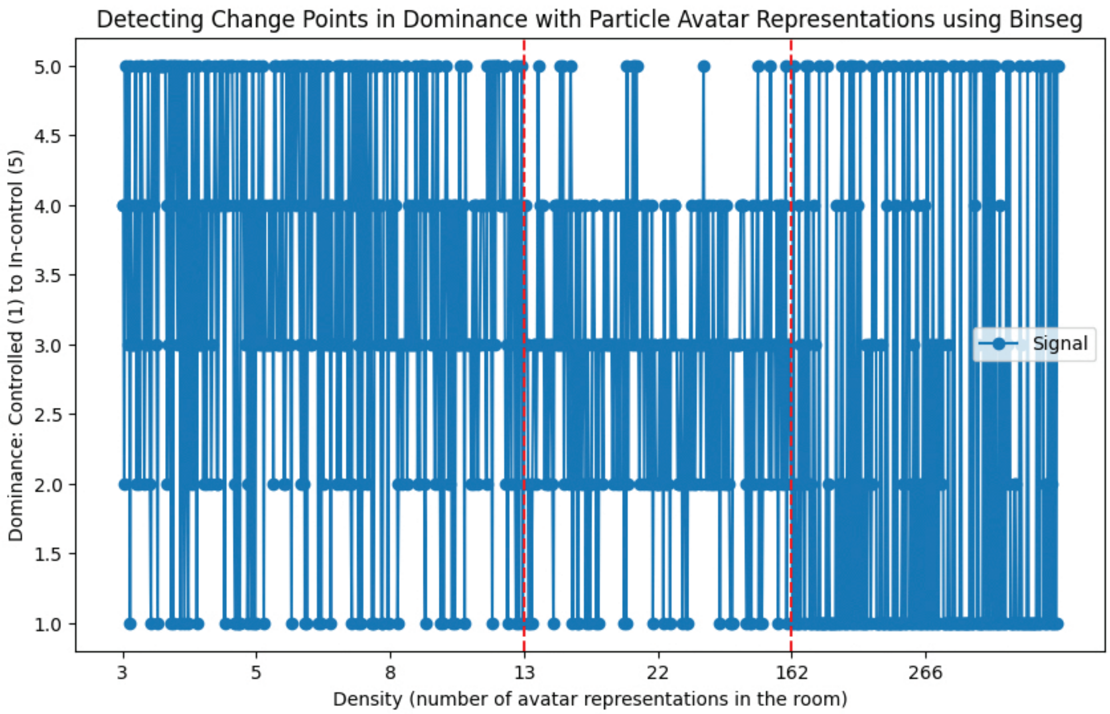

Figure 4.7: Detecting Change Points in Dominance with Particle Avatar Representations using Binseg. The figure illustrates the detection of change points in dominance with particle avatar representations using the Binary Segmentation (Binseg) method. The red dashed lines represent the identified change points at densities 13 and 162.

density	χ^2	*p_value*	**mean particle (rank)**	**mean 2D (rank)**	**mean 3D (rank)**
3	0.500	0.779	1.09 (3rd)	1.11 (1st)	1.09 (2nd)
5	5.591	0.061	1.41 (3rd)	1.46 (2nd)	1.54 (1st)
8	22.521	0.001**	2.01 (2nd)	1.96 (3rd)	2.25 (1st)
13	22.440	0.001**	2.71 (2nd)	2.49 (3rd)	2.84 (1st)
22	6.904	0.032*	3.25 (2nd)	3.25 (3rd)	3.41 (1st)
162	11.405	0.003**	4.82 (2nd)	4.81 (3rd)	4.91 (1st)
266	15.846	0.001**	5.00 (1st)	4.92 (3rd)	4.99 (2nd)

Table 4.13: Friedman Test and Mean Ranks for Crowding Perception

The Table 4.9 includes columns for density, chi-square value (χ^2), p-value, and the mean ranks for each type of stimuli (particle, 2D, and 3D) for crowding perception. Significant p-values (less than 0.05) are marked with an asterisk (*) or double asterisk (**) to indicate their level of significance.

For density 8 and the crowding perception, the non-parametric Friedman test of differences among repeated measures rendered a chi-square value of 22.521, which was significant ($p < 0.001$). To investigate the significant differences identified by the

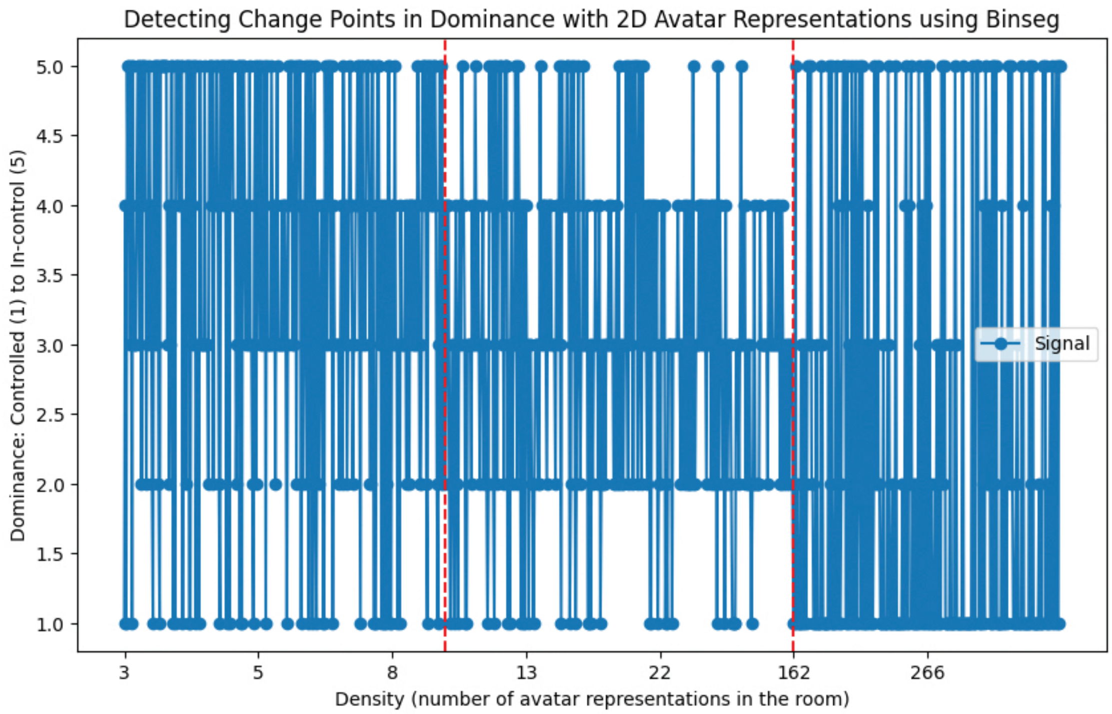

Figure 4.8: Detecting Change Points in Dominance with 2D Avatar Representations using Binseg. The figure illustrates the detection of change points in dominance with 2D avatar representations using the Binary Segmentation (Binseg) method. The red dashed lines represent the identified change points at densities 8 and 162.

Friedman test, post-hoc pairwise comparisons were conducted using the Wilcoxon signed-rank test with Holm correction for multiple comparisons. The following pairs of avatar representations were compared: 2D vs. 3D, 2D vs. Particle, and 3D vs. Particle.

The Wilcoxon signed-rank test results are as follows: For the comparison between 2D and 3D avatars, $W = 849.0$, unadjusted $p = 0.000199$, Holm-adjusted $p = 0.000596$. For the comparison between 2D and Particle avatars, $W = 898.5$, unadjusted $p = 0.415$, Holm-adjusted $p = 0.415$. For the comparison between 3D and Particle avatars, $W = 610.0$, unadjusted $p = 0.000293$, Holm-adjusted $p = 0.000596$. After applying the Holm correction, the differences between the 2D and 3D avatars and between the 3D and Particle avatars were statistically significant at the $\alpha = 0.05$ level. The difference between the 2D and Particle avatars was not statistically significant.

For density 13 and the crowding perception, the non-parametric Friedman test of differences among repeated measures rendered a chi-square value of 22.440, which

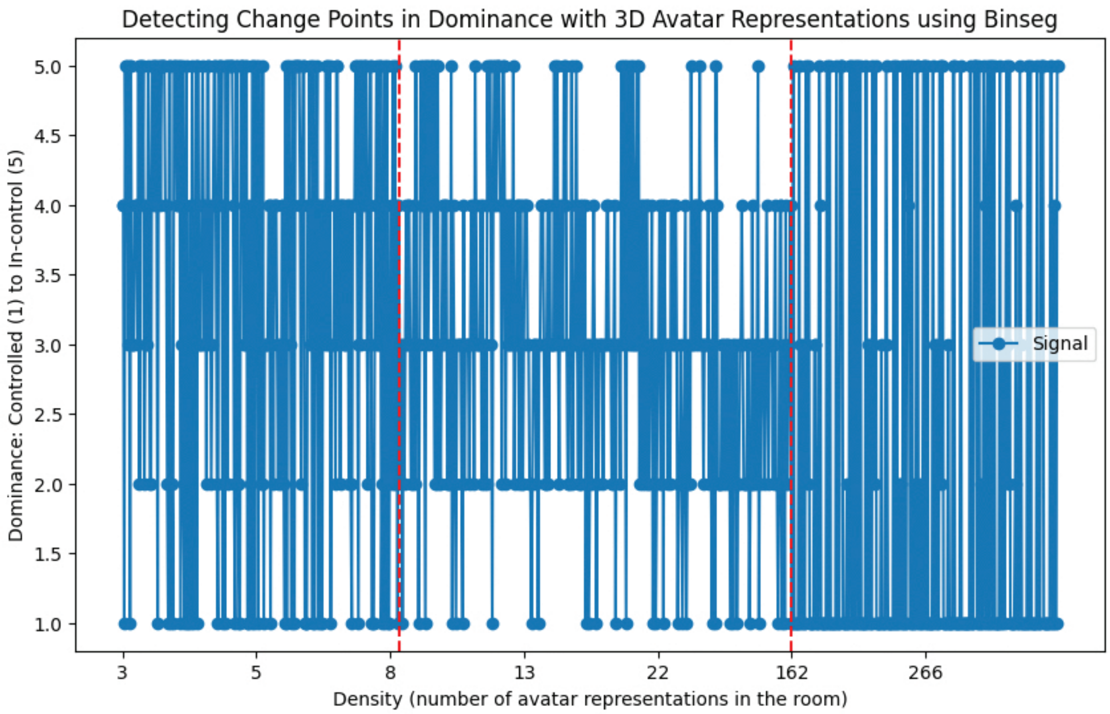

Figure 4.9: Detecting Change Points in Dominance with 3D Avatar Representations using Binseg. The figure illustrates the detection of change points in dominance with 3D avatar representations using the Binary Segmentation (Binseg) method. The red dashed lines represent the identified change points at densities 8 and 162.

was significant ($p < 0.001$). To investigate the significant differences identified by the Friedman test, post-hoc pairwise comparisons were conducted using the Wilcoxon signed-rank test with Holm correction for multiple comparisons. The following pairs of avatar representations were compared: 2D vs. 3D, 2D vs. Particle, and 3D vs. Particle.

The Wilcoxon signed-rank test results are as follows: For the comparison between 2D and Particle avatars, $W = 550.0$, unadjusted $p = 0.002450$, Holm-adjusted $p = 0.004900$. For the comparison between 3D and Particle avatars, $W = 565.5$, unadjusted $p = 0.036992$, Holm-adjusted $p = 0.036992$. For the comparison between 2D and 3D avatars, $W = 438.5$, unadjusted $p = 0.000003$, Holm-adjusted $p = 0.000009$. After applying the Holm correction, the differences between the 2D and Particle avatars, 3D and Particle avatars, and 2D and 3D avatars were statistically significant at the $\alpha = 0.05$ level.

For density 22 and the crowding perception, the non-parametric Friedman test of differences among repeated measures rendered a chi-square value of 6.904, which

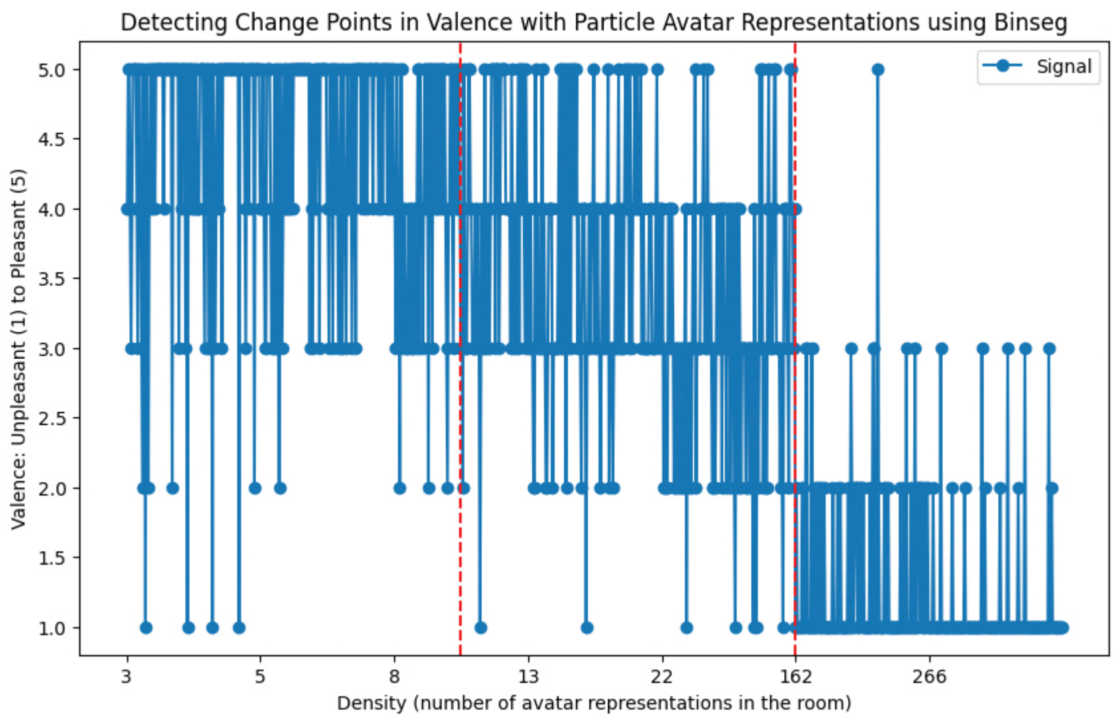

Figure 4.10: Detecting Change Points in Valence with Particle Avatar Representations using Binseg. The figure illustrates the detection of change points in valence with particle avatar representations using the Binary Segmentation (Binseg) method. The red dashed lines represent the identified change points at densities 8 and 162.

was significant ($p = 0.032$). To investigate the significant differences identified by the Friedman test, post-hoc pairwise comparisons were conducted using the Wilcoxon signed-rank test with Holm correction for multiple comparisons. The following pairs of avatar representations were compared: 2D vs. 3D, 2D vs. Particle, and 3D vs. Particle.

The Wilcoxon signed-rank test results are as follows: For the comparison between 2D and Particle avatars, $W = 1134.5$, unadjusted $p = 0.975299$, Holm-adjusted $p = 0.975299$. For the comparison between 3D and Particle avatars, $W = 708.0$, unadjusted $p = 0.022444$, Holm-adjusted $p = 0.046874$. For the comparison between 2D and 3D avatars, $W = 791.5$, unadjusted $p = 0.015625$, Holm-adjusted $p = 0.046874$. After applying the Holm correction, the differences between the 3D and Particle avatars and between the 2D and 3D avatars were statistically significant at the $\alpha = 0.05$ level. The difference between the 2D and Particle avatars was not statistically significant.

For density 162 and the crowding perception, the non-parametric Friedman test

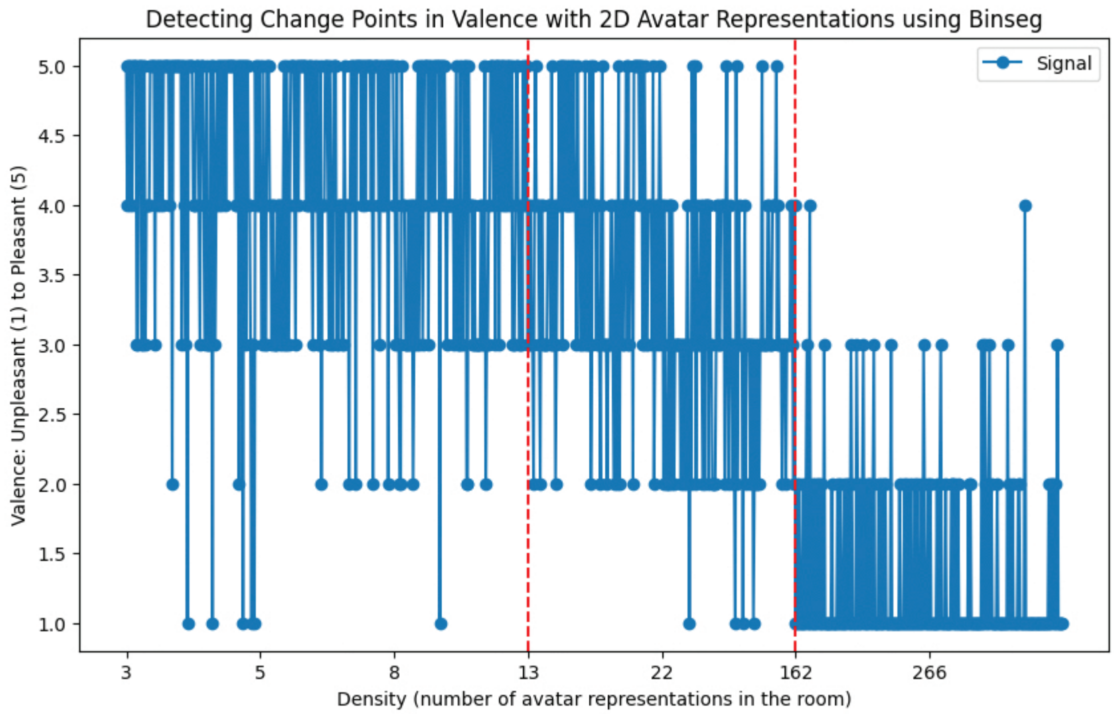

Figure 4.11: Detecting Change Points in Valence with 2D Avatar Representations using Binseg. The figure illustrates the detection of change points in valence with 2D avatar representations using the Binary Segmentation (Binseg) method. The red dashed lines represent the identified change points at densities 13 and 162.

of differences among repeated measures rendered a chi-square value of 11.405, which was significant ($p = 0.003$). To investigate the significant differences identified by the Friedman test, post-hoc pairwise comparisons were conducted using the Wilcoxon signed-rank test with Holm correction for multiple comparisons. The following pairs of avatar representations were compared: 2D vs. 3D, 2D vs. Particle, and 3D vs. Particle.

The Wilcoxon signed-rank test results are as follows: For the comparison between 2D and Particle avatars, $W = 150.0$, unadjusted $p = 0.705457$, Holm-adjusted $p = 0.705457$. For the comparison between 3D and Particle avatars, $W = 81.5$, unadjusted $p = 0.030589$, Holm-adjusted $p = 0.061178$. For the comparison between 2D and 3D avatars, $W = 82.5$, unadjusted $p = 0.018798$, Holm-adjusted $p = 0.056394$. After applying the Holm correction, the differences between the 2D and Particle avatars, 3D and Particle avatars, and 2D and 3D avatars were not statistically significant at the $\alpha = 0.05$ level.

For density 266 and the crowding perception, the non-parametric Friedman test

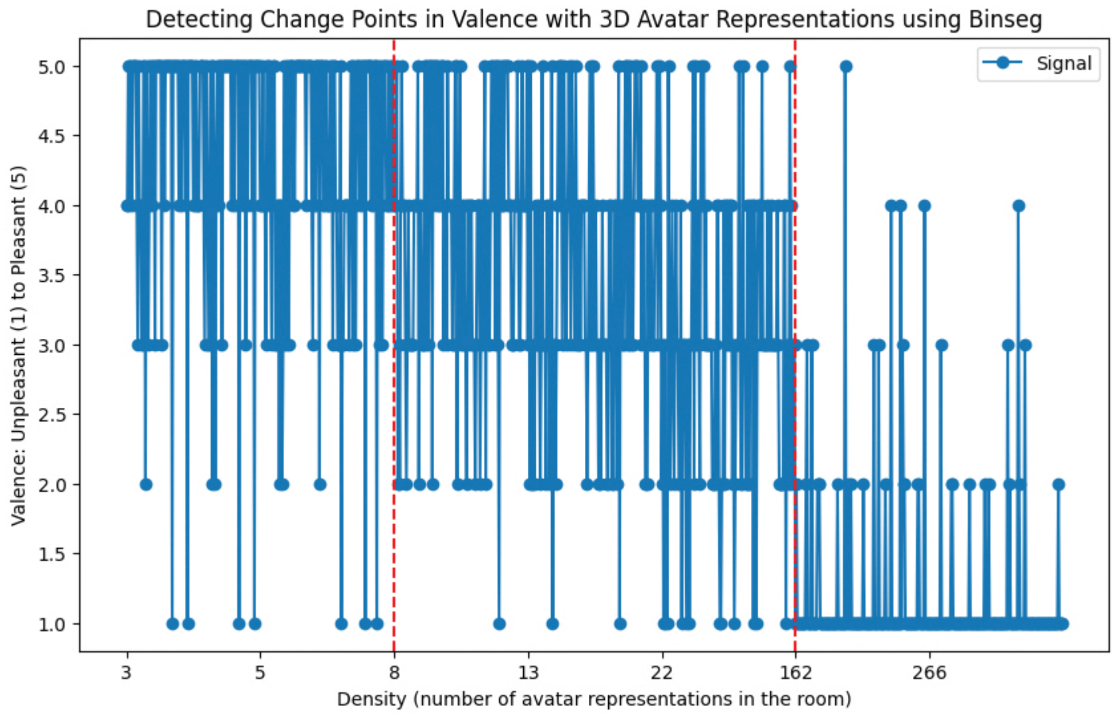

Figure 4.12: Detecting Change Points in Valence with 3D Avatar Representations using Binseg. The figure illustrates the detection of change points in valence with 3D avatar representations using the Binary Segmentation (Binseg) method. The red dashed lines represent the identified change points at densities 8 and 162.

of differences among repeated measures rendered a chi-square value of 15.846, which was significant ($p = 0.001$). To investigate the significant differences identified by the Friedman test, post-hoc pairwise comparisons were conducted using the Wilcoxon signed-rank test with Holm correction for multiple comparisons. The following pairs of avatar representations were compared: 2D vs. 3D, 2D vs. Particle, and 3D vs. Particle.

The Wilcoxon signed-rank test results are as follows: For the comparison between 2D and Particle avatars, $W = 0.0$, unadjusted $p = 0.000911$, Holm-adjusted $p = 0.002733$. For the comparison between 3D and Particle avatars, $W = 0.0$, unadjusted $p = 0.157299$, Holm-adjusted $p = 0.157299$. For the comparison between 2D and 3D avatars, $W = 14.0$, unadjusted $p = 0.012555$, Holm-adjusted $p = 0.025110$. After applying the Holm correction, the differences between the 2D and Particle avatars and between the 2D and 3D avatars were statistically significant at the $\alpha = 0.05$ level. The difference between the 3D and Particle avatars was not statistically significant.

density	χ^2	*p_value*	**mean particle (rank)**	**mean 2D (rank)**	**mean 3D (rank)**
3	3.652	0.161	1.89 (3rd)	2.03 (1st)	1.97 (2nd)
5	7.879	0.019*	1.91 (3rd)	2.13 (1st)	2.01 (2nd)
8	1.709	0.426	2.14 (3rd)	2.23 (1st)	2.22 (2nd)
13	9.324	0.009**	2.56 (2nd)	2.34 (3rd)	2.54 (1st)
22	4.006	0.135	2.74 (3rd)	2.84 (1st)	2.90 (2nd)
162	21.690	0.001**	3.97 (3rd)	3.74 (2nd)	4.04 (1st)
266	19.959	0.001**	4.08 (2nd)	3.80 (3rd)	4.16 (1st)

Table 4.14: Friedman Test and Mean Ranks for Arousal

The Table 4.14 includes columns for density, chi-square value (χ^2), p-value, and the mean ranks for each type of stimuli (particle, 2D, and 3D) for arousal. Significant p-values (less than 0.05) are marked with an asterisk (*) or double asterisk (**) to indicate their level of significance.

For density 5 and the arousal attribute, the non-parametric Friedman test of differences among repeated measures rendered a chi-square value of 7.879, which was significant ($p = 0.019$). To investigate the significant differences identified by the Friedman test, post-hoc pairwise comparisons were conducted using the Wilcoxon signed-rank test with Holm correction for multiple comparisons. The following pairs of avatar representations were compared: 2D vs. 3D, 2D vs. Particle, and 3D vs. Particle.

The Wilcoxon signed-rank test results are as follows: For the comparison between 2D and Particle avatars, $W = 313.0$, unadjusted $p = 0.005085$, Holm-adjusted $p = 0.015256$. For the comparison between 3D and Particle avatars, $W = 298.5$, unadjusted $p = 0.069023$, Holm-adjusted $p = 0.138046$. For the comparison between 2D and 3D avatars, $W = 461.0$, unadjusted $p = 0.240290$, Holm-adjusted $p = 0.240290$. After applying the Holm correction, the difference between the 2D and Particle avatars was statistically significant at the $\alpha = 0.05$ level. The differences between the 3D and Particle avatars and between the 2D and 3D avatars were not statistically significant.

For density 13 and the arousal attribute, the non-parametric Friedman test of differences among repeated measures rendered a chi-square value of 9.324, which was significant ($p = 0.009$). To investigate the significant differences identified by the Friedman test, post-hoc pairwise comparisons were conducted using the Wilcoxon signed-rank test with Holm correction for multiple comparisons. The following pairs of avatar representations were compared: 2D vs. 3D, 2D vs. Particle, and 3D vs. Particle.

The Wilcoxon signed-rank test results are as follows: For the comparison between 2D and Particle avatars, $W = 710.5$, unadjusted $p = 0.007128$, Holm-adjusted $p = 0.021385$. For the comparison between 3D and Particle avatars, $W = 1023.5$, unadjusted $p = 0.905345$, Holm-adjusted $p = 0.905345$. For the comparison between 2D and 3D avatars, $W = 676.0$, unadjusted $p = 0.009268$, Holm-adjusted $p = 0.021385$. After applying the Holm correction, the differences between the 2D and Particle avatars and between the 2D and 3D avatars were statistically significant at the $\alpha = 0.05$ level. The difference between the 3D and Particle avatars was not statistically significant.

For density 162 and the arousal attribute, the non-parametric Friedman test of differences among repeated measures rendered a chi-square value of 21.690, which was significant ($p < 0.001$). To investigate the significant differences identified by the Friedman test, post-hoc pairwise comparisons were conducted using the Wilcoxon signed-rank test with Holm correction for multiple comparisons. The following pairs of avatar representations were compared: 2D vs. 3D, 2D vs. Particle, and 3D vs. Particle.

The Wilcoxon signed-rank test results are as follows: For the comparison between 2D and Particle avatars, $W = 323.0$, unadjusted $p = 0.002106$, Holm-adjusted $p = 0.004211$. For the comparison between 3D and Particle avatars, $W = 513.0$, unadjusted $p = 0.213803$, Holm-adjusted $p = 0.213803$. For the comparison between 2D and 3D avatars, $W = 371.0$, unadjusted $p = 0.000524$, Holm-adjusted $p = 0.001572$. After applying the Holm correction, the differences between the 2D and Particle avatars and between the 2D and 3D avatars were statistically significant at the $\alpha = 0.05$ level. The difference between the 3D and Particle avatars was not statistically significant.

For density 266 and the arousal attribute, the non-parametric Friedman test of differences among repeated measures rendered a chi-square value of 19.959, which was significant ($p < 0.001$). To investigate the significant differences identified by the Friedman test, post-hoc pairwise comparisons were conducted using the Wilcoxon signed-rank test with Holm correction for multiple comparisons. The following pairs of avatar representations were compared: 2D vs. 3D, 2D vs. Particle, and 3D vs. Particle.

The Wilcoxon signed-rank test results are as follows: For the comparison between 2D and Particle avatars, $W = 125.0$, unadjusted $p = 0.002455$, Holm-adjusted $p = 0.004909$. For the comparison between 3D and Particle avatars, $W = 171.5$, unadjusted $p = 0.190307$, Holm-adjusted $p = 0.190307$. For the comparison between 2D and 3D avatars, $W = 112.0$, unadjusted $p = 0.000234$, Holm-adjusted $p = 0.000702$. After applying the Holm correction, the differences between the 2D and Particle avatars and between the 2D and 3D avatars were statistically significant at the $\alpha = 0.05$ level. The difference between the 3D and Particle avatars was not statistically significant.

density	χ^2	*p_value*	**mean particle (rank)**	**mean 2D (rank)**	**mean 3D (rank)**
3	0.245	0.885	3.59 (2nd)	3.69 (1st)	3.56 (3rd)
5	7.224	0.027*	3.57 (2nd)	3.40 (3rd)	3.33 (1st)
8	1.302	0.521	3.38 (1st)	3.23 (3rd)	3.29 (2nd)
13	4.971	0.083	2.93 (3rd)	3.10 (1st)	3.06 (2nd)
22	2.006	0.367	2.75 (2nd)	2.82 (1st)	2.71 (3rd)
162	2.213	0.331	2.07 (3rd)	2.22 (1st)	2.12 (2nd)
266	1.883	0.390	2.09 (2nd)	2.12 (1st)	2.12 (1st)

Table 4.15: Friedman Test and Mean Ranks for Dominance

The Table 4.15 includes columns for density, chi-square value (χ^2), p-value, and the mean ranks for each type of stimuli (particle, 2D, and 3D) for dominance. Significant p-values (less than 0.05) are marked with an asterisk (*) or double asterisk (**) to indicate their level of significance.

For density 5 and the dominance attribute, the non-parametric Friedman test of differences among repeated measures rendered a chi-square value of 7.224, which was significant ($p = 0.027$). To investigate the significant differences identified by the Friedman test, post-hoc pairwise comparisons were conducted using the Wilcoxon signed-rank test with Holm correction for multiple comparisons. The following pairs of avatar representations were compared: 2D vs. 3D, 2D vs. Particle, and 3D vs. Particle.

The Wilcoxon signed-rank test results are as follows: For the comparison between 2D and Particle avatars, $W = 538.0$, unadjusted $p = 0.109003$, Holm-adjusted $p = 0.218007$. For the comparison between 3D and Particle avatars, $W = 668.5$, unadjusted $p = 0.023749$, Holm-adjusted $p = 0.071246$. For the comparison between 2D and 3D avatars, $W = 733.5$, unadjusted $p = 0.446250$, Holm-adjusted $p = 0.446250$. After applying the Holm correction, none of the differences between the pairs of avatars were statistically significant at the $\alpha = 0.05$ level.

density	χ^2	*p_value*	**mean particle (rank)**	**mean 2D (rank)**	**mean 3D (rank)**
3	2.390	0.303	4.46 (1st)	4.35 (2nd)	4.46 (1st)
5	11.522	0.003**	4.44 (1st)	4.23 (3rd)	4.26 (2nd)
8	5.168	0.075	3.98 (1st)	3.95 (2nd)	3.82 (3rd)
13	7.579	0.023*	3.56 (2nd)	3.71 (1st)	3.50 (3rd)
22	3.044	0.218	3.08 (1st)	3.03 (2nd)	2.93 (3rd)
162	14.545	0.001**	1.39 (2nd)	1.39 (2nd)	1.26 (3rd)
266	20.702	0.001**	1.12 (1st)	1.26 (3rd)	1.12 (1st)

Table 4.16: Friedman Test and Mean Ranks for Valence

The Table 4.16 includes columns for density, chi-square value (χ^2), p-value, and the mean ranks for each type of stimuli (particle, 2D, and 3D) for valence. Significant p-values (less than 0.05) are marked with an asterisk (*) or double asterisk (**) to indicate their level of significance.

For density 5 and the valence attribute, the non-parametric Friedman test of differences among repeated measures rendered a chi-square value of 11.522, which was significant ($p = 0.003$). To investigate the significant differences identified by the Friedman test, post-hoc pairwise comparisons were conducted using the Wilcoxon signed-rank test with Holm correction for multiple comparisons. The following pairs of avatar representations were compared: 2D vs. 3D, 2D vs. Particle, and 3D vs. Particle.

The Wilcoxon signed-rank test results are as follows: For the comparison between 2D and Particle avatars, $W = 324.5$, unadjusted $p = 0.002546$, Holm-adjusted $p = 0.007638$. For the comparison between 3D and Particle avatars, $W = 283.0$, unadjusted $p = 0.003930$, Holm-adjusted $p = 0.007859$. For the comparison between 2D and 3D avatars, $W = 498.0$, unadjusted $p = 0.456911$, Holm-adjusted $p = 0.456911$. After applying the Holm correction, the differences between the 2D and Particle avatars and between the 3D and Particle avatars were statistically significant at the $\alpha = 0.05$ level. The difference between the 2D and 3D avatars was not statistically significant.

For density 13 and the valence attribute, the non-parametric Friedman test of differences among repeated measures rendered a chi-square value of 7.579, which was significant ($p = 0.023$). To investigate the significant differences identified by the Friedman test, post-hoc pairwise comparisons were conducted using the Wilcoxon signed-rank test with Holm correction for multiple comparisons. The following pairs of avatar representations were compared: 2D vs. 3D, 2D vs. Particle, and 3D vs. Particle.

The Wilcoxon signed-rank test results are as follows: For the comparison between 2D and Particle avatars, $W = 879.0$, unadjusted $p = 0.034979$, Holm-adjusted $p = 0.069959$. For the comparison between 3D and Particle avatars, $W = 908.0$, unadjusted $p = 0.454347$, Holm-adjusted $p = 0.454347$. For the comparison between 2D and 3D avatars, $W = 655.5$, unadjusted $p = 0.001820$, Holm-adjusted $p = 0.005460$. After applying the Holm correction, the difference between the 2D and 3D avatars was statistically significant at the $\alpha = 0.05$ level. The differences between the 2D and Particle avatars and between the 3D and Particle avatars were not statistically significant.

For density 162 and the valence attribute, the non-parametric Friedman test of differences among repeated measures rendered a chi-square value of 14.545, which was significant ($p < 0.001$). To investigate the significant differences identified by the Friedman test, post-hoc pairwise comparisons were conducted using the Wilcoxon signed-rank test with Holm correction for multiple comparisons. The following pairs of avatar representations were compared: 2D vs. 3D, 2D vs. Particle, and 3D vs. Particle.

The Wilcoxon signed-rank test results are as follows: For the comparison between 2D and Particle avatars, $W = 325.0$, unadjusted $p = 0.891365$, Holm-adjusted $p = 0.891365$. For the comparison between 3D and Particle avatars, $W = 203.0$, unadjusted $p = 0.031125$, Holm-adjusted $p = 0.062250$. For the comparison between 2D and 3D avatars, $W = 247.5$, unadjusted $p = 0.019634$, Holm-adjusted $p = 0.058901$. After applying the Holm correction, the differences between the pairs of avatars were not statistically significant at the $\alpha = 0.05$ level.

For density 266 and the valence attribute, the non-parametric Friedman test of differences among repeated measures rendered a chi-square value of 20.702, which was significant ($p < 0.001$). To investigate the significant differences identified by the Friedman test, post-hoc pairwise comparisons were conducted using the Wilcoxon signed-rank test with Holm correction for multiple comparisons. The following pairs of avatar representations were compared: 2D vs. 3D, 2D vs. Particle, and 3D vs. Particle.

The Wilcoxon signed-rank test results are as follows: For the comparison between 2D and Particle avatars, $W = 50.0$, unadjusted $p = 0.000413$, Holm-adjusted $p = 0.001238$. For the comparison between 3D and Particle avatars, $W = 45.0$, unadjusted $p = 0.970566$, Holm-adjusted $p = 0.970566$. For the comparison between 2D and 3D avatars, $W = 43.5$, unadjusted $p = 0.001057$, Holm-adjusted $p = 0.002113$. After applying the Holm correction, the differences between the 2D and Particle avatars and between the 2D and 3D avatars were statistically significant at the $\alpha = 0.05$ level. The difference between the 3D and Particle avatars was not statistically significant.

Exploring relationships

In Table 4.17, the intercorrelations between spatial crowding, arousal, valence, and dominance were showed. Spatial crowding was found to have a strong negative correlation with valence ($r = -0.72$), indicating that as perceived crowding increases, feelings of valence decrease. Conversely, spatial crowding had a positive correlation with arousal ($r = 0.24$), suggesting that higher levels of crowding are associated with increased arousal. The correlation between spatial crowding and dominance was weakly negative ($r = -0.05$), implying a negligible relationship. Arousal was positively and weakly correlated with dominance ($r = 0.14$) and weakly negatively correlated with valence ($r = -0.18$), indicating that higher arousal levels are slightly associated with lower valence and higher dominance. Valence showed a weak positive correlation with dominance ($r = 0.22$), suggesting that higher valence is moderately associated with higher dominance.

	1	2	3
1. Spatial Crowding			
2. Arousal (SAM)	0.24		
3. Valence (SAM)	-0.72	-0.18	
4. Dominance (SAM)	-0.05	0.14	0.22

Table 4.17: Intercorrelations of Subjective Measures. Table presents the intercorrelations between subjective measures, including spatial crowding, arousal (SAM), valence (SAM), and dominance (SAM), showing the relationships between these variables.

4.2 Main Study

To check the replicability of our pilot results and extend them, a main study was conducted. The main study ran from February 12, 2024, to March 16, 2024. Conducted online using the Qualtrics platform, the study included 48 participants. Using a within-subjects design, participants were exposed to experimental conditions varying in avatar representation (particle, 2D, and 3D) and density. The valid densities evaluated in the pilot study were 3, 5, 8, 13, 22, 36, 60, 98, 162, and 266, resulting in 30 different stimuli and images. Unlike the pilot study, each stimulus in the main study was shown to participants three times.

4.2.1 Demographics Report

Participants

A total of 48 participants completed the survey. The sample consisted of 33 women (68.8%) and 15 men (31.3%).

Assigned Sex at Birth	n	%
Woman	33	68.8
Man	15	31.3

Table 4.18: Assigned Sex at Birth of Participants (N = 48)

Education Level

Participants reported various levels of education. The largest groups of participants held a 4-year degree (33.3%) or a professional degree (33.3%). Additionally, 12.5% of participants were currently working on a degree, 10.4% had completed a doctorate, 8.3% were high school graduates, and 2.1% held a 2-year degree.

Education Level	n	%
2-year degree	1	2.1
4-year degree	16	33.3
Degree in progress (currently working on a degree)	6	12.5
Doctorate	5	10.4
High school graduate	4	8.3
Professional degree	16	33.3

Table 4.19: Highest Completed Education Level of Participants (N = 48)

Design Experience

When asked about their design experience, 44 participants (91.7%) reported that they do not consider themselves designers, while 4 participants (8.3%) identified as designers.

Do you consider yourself a designer?	n	%
No	44	91.7
Yes	4	8.3

Table 4.20: Design Experience of Participants (N = 48)

4.2.2 Consistency Analysis for Repeated Stimuli

To address the research question of how users respond to viewing the same image multiple times, I investigated the differences and similarities in their reactions. Our primary research question (RQ1) is: **How consistent are users' crowding perceptions and emotional responses (valence, arousal, dominance) when exposed to varying crowd densities and different avatar representations (particle, 2D, and 3D) multiple times?**

I tested if the data was distributed normally or not by conducting the Shapiro-Wilk test prior to making comparisons between stimuli. The results of the test showed that for all stimuli, p-value < 0.001, indicating that none of these followed a normal distribution. This is not surprising since such samples taken from Likert scales usually do not follow normal distributions. It should be noted that Likert responses are ordinal and although they are often subjected to parametric tests that presume interval-level measurement, this method is frequently inappropriate [37]. Therefore, it was necessary to utilize non-parametric tests for further analysis.

In order for me to learn about any trends or inconsistencies in their responses, I conducted the Friedman test. This method allowed me to determine if there were any notable differences in the consistency of their reactions.

The Friedman test was then performed to evaluate the differences in participants' responses across various conditions. This test is a non-parametric statistical test used to detect differences in treatments across multiple test attempts. Specifically, it checks if there is a significant difference in the rankings of responses when participants are shown different stimuli. The rationale is that if the stimuli have an effect on the participants' ratings, the rankings will differ significantly across conditions. The Friedman test helps verify these differences, ensuring that any observed differences in rankings are not due to random variation.

The heatmap in Figure 4.13 illustrates the p-values derived from the Friedman test across various stimuli, densities, and attributes. The p-values are color-coded, with lower p-values (indicating statistically significant differences) represented in shades of

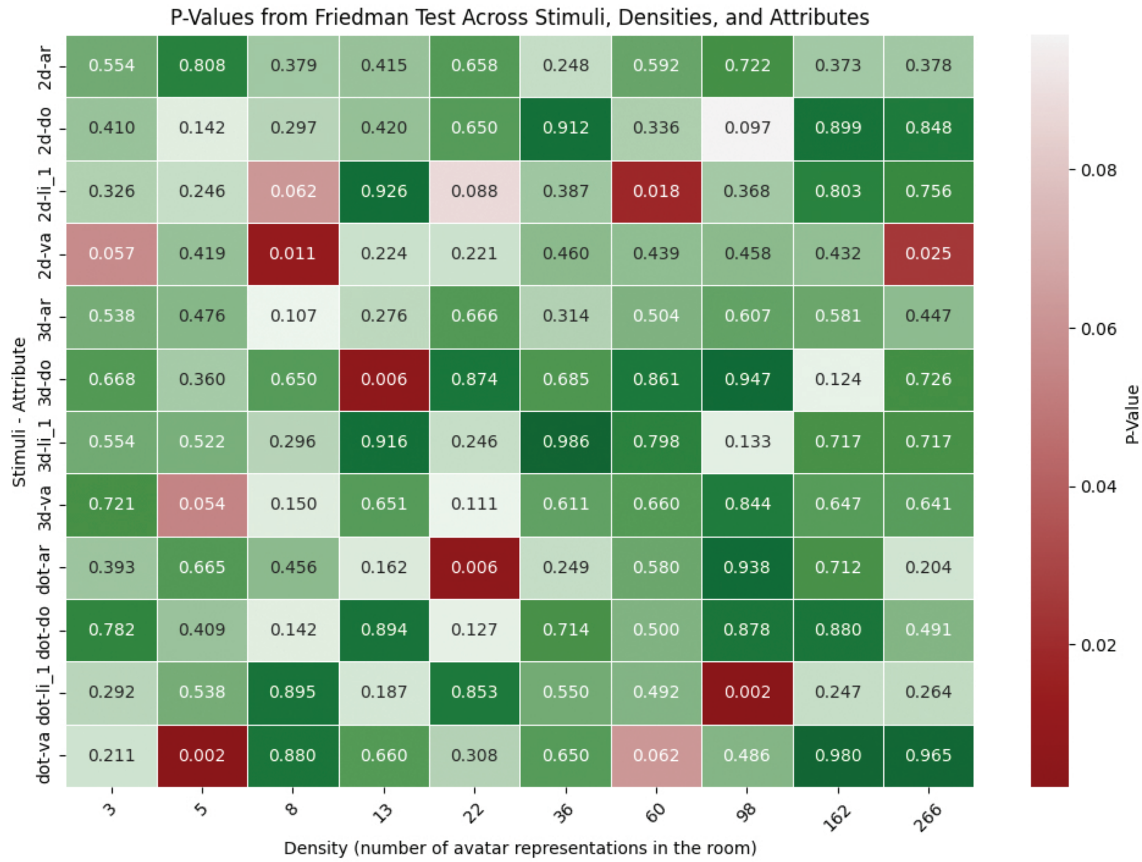

Figure 4.13: Friedman Test across stimuli. The image illustrates the p-values derived from the Friedman test across various stimuli, densities, and attributes. The p-values are color-coded, with lower p-values (indicating statistically significant differences) represented in shades of red and higher p-values in shades of green. The y-axis labels in the heatmap represent combinations of stimuli types (2D, 3D, Particle) and attributes (Valence, Crowding, Arousal, Dominance), indicating, for example, "$2d - va$" for 2D valence, "$3d - li_1$" for 3D crowding, and "$dot - ar$" for particle arousal, with each label describing the specific attribute being tested across varying densities shown on the x-axis.

red and higher p-values in shades of green.

The Friedman test identifies statistically significant differences for particle avatar representations. For valence at density 5, the non-parametric Friedman test of differences among repeated measures rendered a chi-square value of 12.685, which was significant ($p = 0.002$). For arousal at density 22, the non-parametric Friedman test of differences among repeated measures rendered a chi-square value of 10.196, which was significant ($p = 0.006$). For crowding perception at density 98, the non-parametric Friedman test of differences among repeated measures rendered a chi-square value of

12.043, which was significant ($p = 0.002$).

The Friedman test identified statistically significant differences for 2D avatar representations. For valence at density 8, the non-parametric Friedman test of differences among repeated measures rendered a chi-square value of 9.073, which was significant ($p = 0.011$). For crowding perception at density 60, the non-parametric Friedman test of differences among repeated measures rendered a chi-square value of 8, which was significant ($p = 0.018$). For valence at density 266, the non-parametric Friedman test of differences among repeated measures rendered a chi-square value of 7.396, which was significant ($p = 0.025$).

The Friedman test identified a single statistically significant difference for 3D avatar representations. For dominance at density 13, the non-parametric Friedman test of differences among repeated measures rendered a chi-square value of 10.393, which was significant ($p = 0.006$).

Post-hoc pairwise comparisons were not conducted as we were comparing the same image three times and all the images were randomly seen by the participants, without any specific order or point in time to be analyzed.

Creating composite scores from multiple variables

To create a composite variable, again I used the Intraclass Coefficient class (C, k) [80, 80].

Under the assumption of randomly selected raters, the ICC for interrater consistency (C, k) assesses the reliability of the observed differences between subjects by accounting for the variance due to subjects while excluding the rater effects. This means it measures how consistently subjects are rated across different raters, and for average ratings, it divides the error variance by the number of raters to reduce rater-related errors. Essentially, ICC(C, k) for interrater consistency focuses on excluding the variability introduced by different raters rating the same image differently at different times. It aims to measure the consistency of the observed differences between subjects while accounting for the average ratings across multiple raters, thereby reducing the impact of rater-related errors or inconsistencies.

The intraclass correlation coefficients (ICCs) for different densities and attributes were calculated across 2D, 3D, and Dot stimuli types. The results in Table 4.21 indicated high reliability across most conditions. Regarding the 2D stimuli type, the ICC values ranged from 0.674 to 0.927 for crowding perception, indicating moderate to high reliability. For valence, the ICC values ranged from 0.810 to 0.928, indicating high reliability. For arousal, the ICC values ranged from 0.784 to 0.944, indicating high reliability. For dominance, the ICC values ranged from 0.812 to 0.923, indicating high reliability.

Regarding the 3D stimuli type, the ICC values ranged from 0.616 to 0.892, indicating moderate to high reliability for crowding perception. For valence, the ICC values ranged from 0.762 to 0.932, indicating high reliability. For arousal, the ICC values ranged from 0.793 to 0.944, indicating high reliability. For dominance, the ICC values ranged from 0.795 to 0.955, indicating high reliability.

Regarding the particle stimuli type, the ICC values ranged from 0.758 to 0.902, indicating high reliability for crowding perception. For valence, the ICC values ranged from 0.762 to 0.889, indicating high reliability. For arousal, the ICC values ranged from 0.781 to 0.897, indicating high reliability. For dominance, the ICC values ranged from 0.812 to 0.931, indicating high reliability.

Compared to the pilot study, it seems that adding one more image in the main study increased reliability, resulting in higher ICC scores. Based on these results, a composite variable can be created by combining the two ratings for each stimulus type and density into a single variable.

4.2.3 Descriptive analysis and thresholds

Stimuli	Crowding Thresholds	Valence Thresholds	Arousal Thresholds	Dominance Thresholds
Particle	[8, 13, 60]	[8, 36, 98]	[8, 36, 60]	[8, 60]
2D	[8, 22, 60]	[8, 22, 98]	[8, 60, 162]	[8, 98]
3D	[8, 36, 98]	[8, 36, 98]	[13, 36, 98]	[13, 98]

Table 4.22: Change Point Detection using Binary Segmentation for Different Stimuli Types and Attributes

Understanding the point at which density becomes crowding and how it impacts emotional responses is crucial for various applications, especially in the design of virtual environments. In this section, I will explore the methods to determine these threshold densities. Our research question (RQ2) is: **Is there a threshold density at which users' emotional responses significantly change?**

To assess thresholds, the Changepoint Detection method will be used. The Binary Segmentation (Binseg) method allows for precise identification of shifts in ratings, providing insights into how perception varies with different avatar densities in 2D, 3D, and Dot virtual environments. Change Point is a moment of time when a time series changes its behaviour [2]. An overview of the change points can be seen in Table 4.22, and the next sections will report on it in more detail.

Analysis of Crowding Perception Across Densities

	Density (number of avatar representations in the room)									
Stimuli	**3**	**5**	**8**	**13**	**22**	**36**	**60**	**98**	**162**	**266**
Particle	1.24 (0.47)	1.58 (0.78)	2.27 (0.86)	2.80 (0.87)	3.36 (0.93)	3.65 (0.90)	4.18 (0.85)	4.71 (0.60)	4.90 (0.32)	4.95 (0.27)
2D	1.21 (0.55)	1.60 (0.82)	2.06 (0.94)	2.70 (1.01)	3.22 (0.99)	3.69 (0.94)	4.15 (0.79)	4.44 (0.76)	4.72 (0.54)	4.86 (0.39)
3D	1.21 (0.47)	1.56 (0.74)	2.06 (0.84)	2.72 (0.89)	3.24 (0.97)	3.67 (0.93)	4.07 (0.90)	4.45 (0.77)	4.77 (0.56)	4.93 (0.28)

Table 4.23: Descriptive statistics for crowding perception across different densities, with mean values and standard deviations in parentheses. The table from the main study indicates that mean values generally increase with density for 2D, 3D, and particle stimuli, with standard deviations reflecting varying degrees of variability.

Participants report a greater crowding perception in environments with higher densities, as this trend holds true across all three categories of stimuli (2D, 3D, and particles). Furthermore, it appears that participants' impressions of crowding become more consistent at higher densities, as standard deviations tend to decrease as density increases. There is a clear agreement across individuals, particularly when it comes to 3D stimuli, since the variety in responses (standard deviation) decreases with increasing densities.

The descriptive statistics for crowding perception across different densities and stimuli types are summarized in Table 4.23. For the 2D stimuli, the mean crowding perception values ranged from 1.21 ($SD = 0.55$) at a density of 3 to 4.86 ($SD = 0.39$) at a density of 266. For the 3D stimuli, the mean values ranged from 1.21 ($SD = 0.47$) at a density of 3 to 4.93 ($SD = 0.28$) at a density of 266. Lastly, for the particle stimuli, the mean values ranged from 1.24 ($SD = 0.47$) at a density of 3 to 4.95 ($SD = 0.27$) at a density of 266.

The Figure 4.14 illustrates the detection of change points in crowding perception with particle, 2D and 3D avatar representations using the Binary Segmentation (Binseg) method. The x-axis represents the density, while the y-axis represents the crowding perception, ranging from Not Crowded at All to Way Too Crowded (1-5). The blue points indicate the signal, with red dashed lines marking the identified change points.

For particle avatar representations, change points were detected at the densities 8, 13, 60. The data shows that between densities 3 and 8, most participants reported a

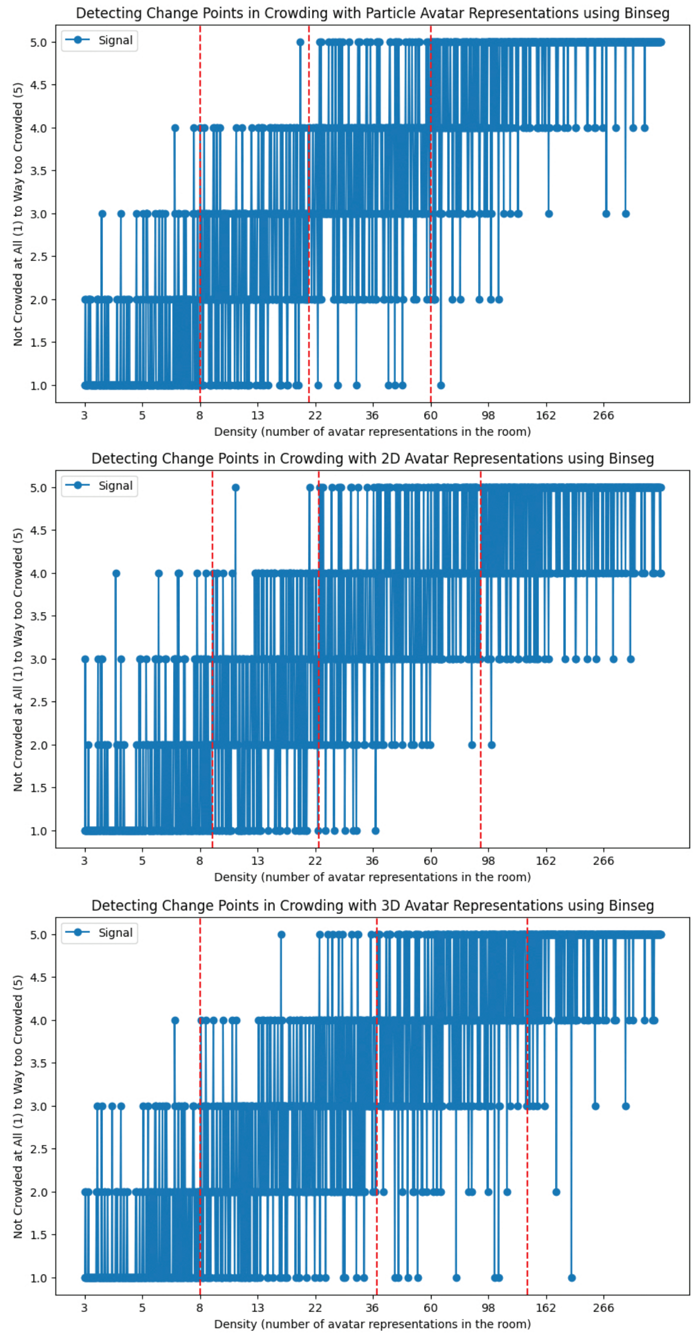

Figure 4.14: Detecting Change Points in Crowding Perception with Particle, 2D and 3D Avatar Representations using Binseg. The figure illustrates the detection of change points in crowding perception with avatar representations using the Binary Segmentation (Binseg) method. The red dashed lines represent the identified change points.

perception of ”Not crowded at all” to ”Not crowded.” Between densities 8 and 13, the concentration of responses shifted predominantly to ”Not crowded.” Between densities 22 and 60, participants generally reported the crowding level as ”Comfortably Occupied.” Above density 60, there was a noticeable shift, with many participants rating the environment as ”Somewhat crowded” or ”Way too Crowded.”

For 2D avatar representations, change points were detected at the densities 8, 22, 60. The data shows that between densities 3 and 8, most participants reported a perception of ”Not crowded at all” to ”Not crowded.” Between densities 8 and 22, the concentration of responses shifted predominantly to ”Not crowded” to ”Comfortably Occupied.” Between densities 22 and 60, participants generally reported the crowding level as ”Comfortably Occupied” to ”Somewhat crowded.” Above density 60, there was a noticeable shift, with many participants rating the environment as ”Somewhat crowded” or ”Way too Crowded.”

For 3D avatar representations, change points were detected at densities 8, 36, and 98. The data show that between densities 3 and 8, most participants reported a perception of ”Not crowded at all” to ”Not crowded.” Between densities 8 and 36, the concentration of responses was more scattered between ”Not crowded” and ”Comfortably Occupied.” Similarly, between densities 36 and 98, responses were scattered between ”Comfortably Occupied” and ”Way too Crowded.” Above density 98, many participants rated the environment as ”Way too Crowded.”

Analysis of Arousal Across Densities

Participant responses show that arousal positively correlated with the number of avatars present in the space. The standard deviation for 3D stimuli decreased from a density of 5 to 36, then increased again at a density of 60. At densities below approximately 36, participants reported an arousal level of 3 or below, suggesting sensations of inactivity such as lethargy or monotony. However, when the densities reached 60 or higher, participants exhibited higher arousal levels, indicating more intense feelings of activity, such as excitement or alertness.

	Density (number of avatar representations in the room)									
Stimuli	**3**	**5**	**8**	**13**	**22**	**36**	**60**	**98**	**162**	**266**
2D	1.79 (1.14)	1.85 (1.04)	2.06 (1.05)	2.31 (0.90)	2.63 (0.92)	2.83 (1.05)	3.15 (1.19)	3.51 (1.23)	3.64 (1.33)	3.83 (1.40)
3D	1.86 (1.22)	1.86 (1.01)	2.02 (0.92)	2.37 (0.99)	2.71 (0.93)	3.10 (1.07)	3.28 (1.17)	3.56 (1.28)	3.87 (1.27)	4.11 (1.35)
Dot	1.76 (1.07)	1.94 (1.09)	2.10 (0.95)	2.42 (0.93)	2.80 (1.07)	2.93 (1.03)	3.35 (1.14)	3.73 (1.31)	3.85 (1.40)	4.07 (1.41)

Table 4.24: Descriptive statistics for arousal across different densities, with mean values and standard deviations in parentheses. The table from the main study indicates that mean values generally increase with density for 2D, 3D, and particle stimuli, with standard deviations reflecting varying degrees of variability.

The descriptive statistics for arousal across different densities and stimuli types are summarized in Table 4.24. For the 2D stimuli, the mean arousal values ranged from 1.79 ($SD = 1.14$) at a density of 3 to 3.83 ($SD = 1.40$) at a density of 266. For the 3D stimuli, the mean values ranged from 1.86 ($SD = 1.22$) at a density of 3 to 4.11 ($SD = 1.35$) at a density of 266. Lastly, for the particle stimuli, the mean values ranged from 1.76 ($SD = 1.07$) at a density of 3 to 4.07 ($SD = 1.41$) at a density of 266.

The Figure 4.15 illustrates the detection of change points in arousal with particle, 2D and 3D avatar representations using the Binary Segmentation (Binseg) method. The x-axis represents the density, while the y-axis represents the arousal scores, ranging from Calm to Excited (1-5). The blue points indicate the signal, with red dashed lines marking the identified change points.

For particle avatar representations, change points were detected at the densities 8, 36, 60. The data shows that between densities 3 and 8, most participants reported arousal as "Calm." Between densities 8 and 36, the concentration of responses shifted predominantly to "Dull" and "Neutral." Between densities 36 and 60, participants generally reported the arousal level as "Wide-awake." Above density 60, responses were more scattered between "Wide-awake" or "Excited."

For 2D avatar representations, change points were detected at densities 8, 60, and 162. The data show that between densities 3 and 8, most participants reported arousal of "Calm" to "Dull." Between densities 13 and 60, the concentration of responses shifted predominantly to "Dull" to "Neutral." Between densities 60 and 162, participants generally reported the arousal level as "Neutral" to "Wide-awake." Above density 162, there was a noticeable shift, with many participants rating the environment as "Excited."

For 3D avatar representations, change points were detected at densities 13, 36, and 98. The data show that between densities 3 and 13, most participants reported arousal of "Calm" to "Dull." Between densities 13 and 36, the concentration of re-

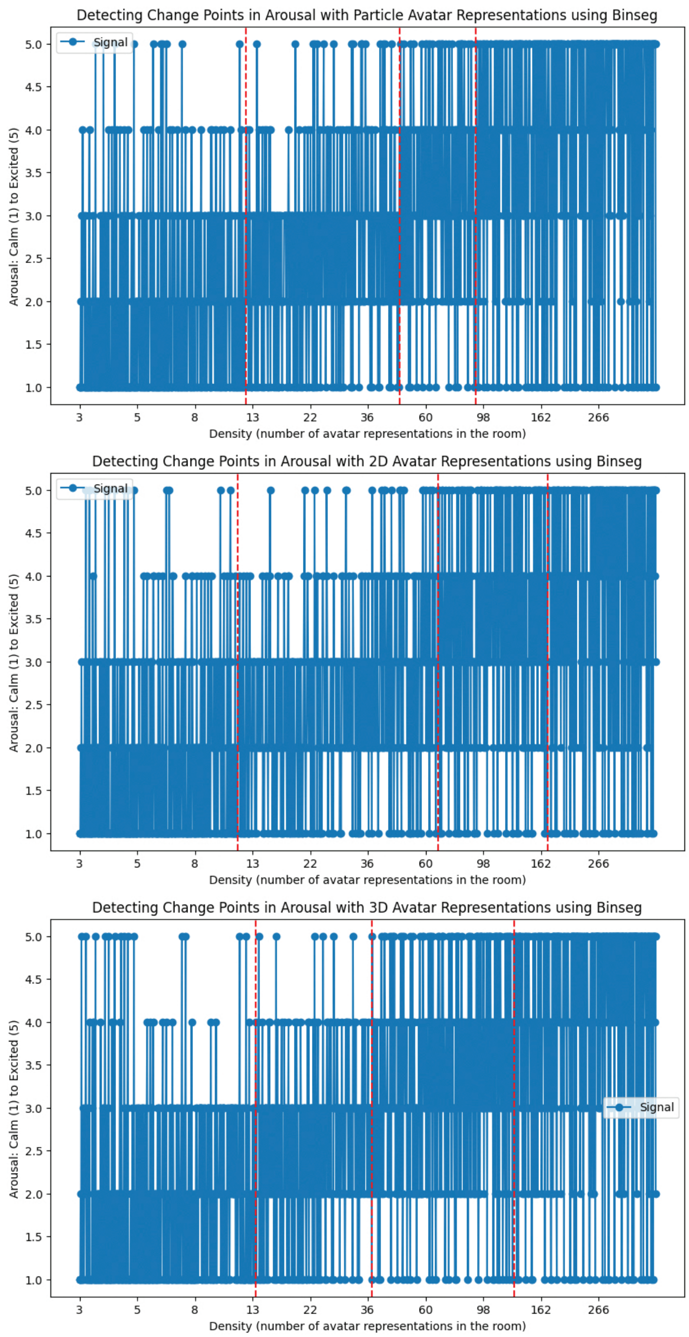

Figure 4.15: Detecting Change Points in Arousal with Particle, 2D and 3D Avatar Representations using Binseg. The figure illustrates the detection of change points in arousal with avatar representations using the Binary Segmentation (Binseg) method. The red dashed lines represent the identified change points.

sponses shifted predominantly to "Dull" to "Neutral." Between densities 36 and 98, participants generally reported the arousal level as "Neutral" to "Wide-awake." Above density 162, many participants rated arousal level as "Excited."

	Density (number of avatar representations in the room)									
Stimuli	**3**	**5**	**8**	**13**	**22**	**36**	**60**	**98**	**162**	**266**
Particle	3.39 (1.56)	3.40 (1.46)	3.19 (1.31)	2.99 (1.04)	2.94 (1.07)	2.73 (1.03)	2.58 (1.21)	2.43 (1.46)	2.31 (1.59)	2.35 (1.69)
2D	3.44 (1.60)	3.35 (1.44)	3.40 (1.31)	3.10 (1.13)	2.90 (1.09)	2.73 (1.00)	2.72 (1.17)	2.47 (1.29)	2.29 (1.38)	2.34 (1.55)
3D	3.32 (1.63)	3.29 (1.47)	3.19 (1.27)	2.96 (1.11)	2.79 (0.92)	2.58 (1.05)	2.58 (1.19)	2.41 (1.30)	2.38 (1.54)	2.24 (1.64)

Table 4.25: Descriptive statistics for dominance across different densities, with mean values and standard deviations in parentheses. The table from the main study indicates that mean values generally decrease with density for 2D, 3D, and particle stimuli, with standard deviations reflecting varying degrees of variability.

Analysis of Dominance Across Densities

Dominance exhibits a different behavior compared to the other subjective measures. While all avatar types show a progressive reduction in dominance as density increases, the changes are not as consistent as for the other attributes. For this reason, dominance has mainly two change points. For 3D stimuli, the mean dominance perception is very similar at densities 36 and 60, and again at densities 98 and 162. A similar pattern occurs for 2D stimuli at densities 36 and 60. The feeling of dominance does not decrease as steadily for particle stimuli and does not always drop. Overall, as density increases for all avatar types, the sense of empowerment and being in control diminishes, although the severity varies. The standard deviation fluctuates somewhat but remains relatively constant.

The descriptive statistics for dominance across different densities and stimuli types are summarized in Table 4.25. For the 2D stimuli, the mean dominance values ranged from 3.44 ($SD = 1.60$) at a density of 3 to 2.34 ($SD = 1.55$) at a density of 266. For the 3D stimuli, the mean values ranged from 3.32 ($SD = 1.63$) at a density of 3 to 2.24 ($SD = 1.64$) at a density of 266. Lastly, for the particle stimuli, the mean values ranged from 3.39 ($SD = 1.56$) at a density of 3 to 2.35 ($SD = 1.69$) at a density of 266.

The Figure 4.16 illustrates the detection of change points in dominance with particle, 2D and 3D avatar representations using the Binary Segmentation (Binseg) method. The x-axis represents the density, while the y-axis represents the dominance scores, ranging from Controlled to In-control (1-5). The blue points indicate the signal, with red dashed lines marking the identified change points.

For particle avatar representations, change points were detected at the densities 8 and 60. The data shows that between densities 3 and 8, participants reported feelings

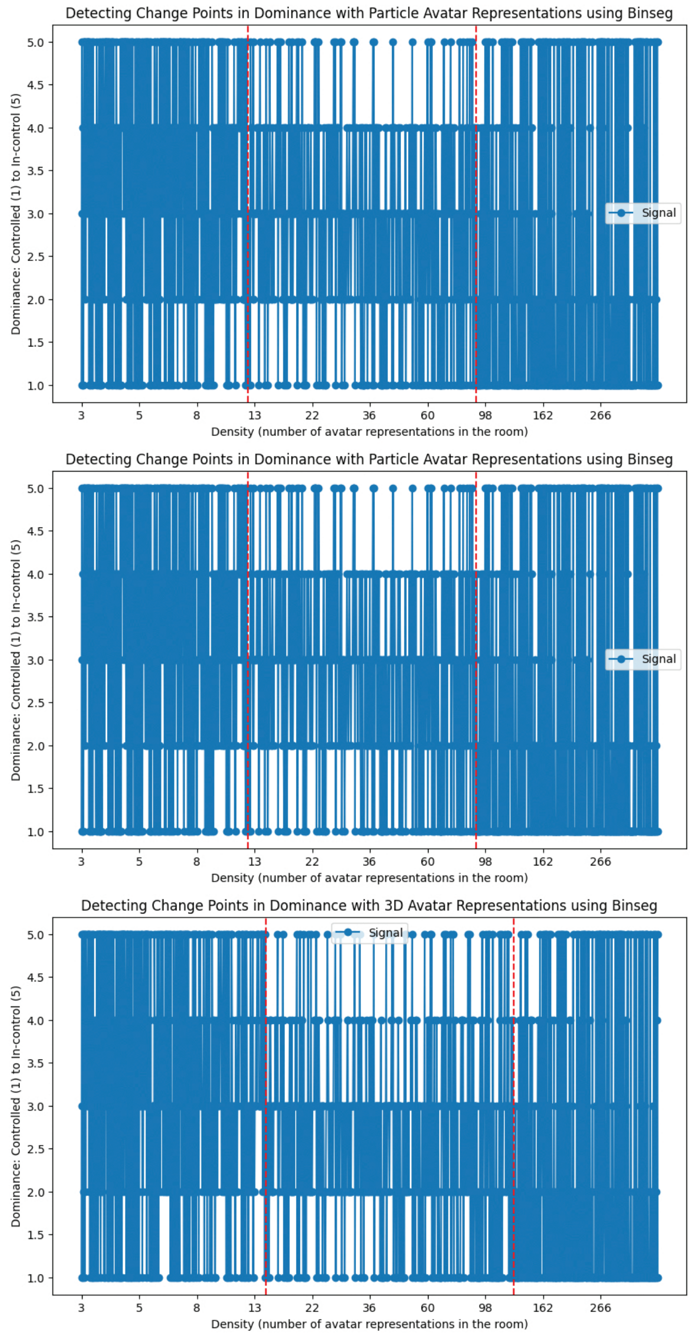

Figure 4.16: Detecting Change Points in Dominance with Particle, 2D and 3D Avatar Representations using Binseg. The figure illustrates the detection of change points in dominance with avatar representations using the Binary Segmentation (Binseg) method. The red dashed lines represent the identified change points.

ranging from "In-control" to "Neutral, indicating higher levels of dominance. Between densities 13 and 60, the responses showed a wider distribution, with participants reporting from "Powerful" to "Powerlessness." This range suggests a decreasing sense of dominance as density grows within this interval. Above density 98, the perceived dominance remained diverse but indicated a slight tendency towards "Powerlessness" to "Controlled."

For 2D avatar representations, change points were detected at densities 8 and 98. The data shows that between densities 3 and 8, participants reported feelings ranging from "In-control" to "Neutral, indicating higher levels of dominance. Between densities 13 and 98, the responses showed a wider distribution, with participants reporting from "Powerful" to "Powerlessness." This range suggests a decreasing sense of dominance as density grows within this interval. Above density 98, the perceived dominance remained diverse but indicated a slight tendency towards "Powerlessness" to "Controlled."

For 3D avatar representations, change points were detected at densities 13 and 98. The data shows that between densities 3 and 13, participants reported feelings ranging from "In-control" to "Neutral, indicating higher levels of dominance. Between densities 13 and 98, the responses showed a wider distribution, with participants reporting from "Neutral" to "Powerlessness." This range suggests a decreasing sense of dominance as density grows within this interval. Above density 98, the perceived dominance remained diverse but indicated a slight tendency towards "Powerlessness" to "Controlled."

	Density (number of avatar representations in the room)									
Stimuli	**3**	**5**	**8**	**13**	**22**	**36**	**60**	**98**	**162**	**266**
Particle	4.36 (0.84)	4.21 (0.89)	3.74 (0.93)	3.26 (0.95)	2.88 (0.99)	2.57 (0.92)	2.24 (0.96)	1.75 (0.96)	1.49 (0.84)	1.30 (0.81)
2D	4.38 (0.90)	4.01 (0.94)	3.76 (0.98)	3.26 (0.99)	2.96 (1.00)	2.61 (0.91)	2.31 (1.00)	2.02 (0.98)	1.72 (0.93)	1.43 (0.78)
3D	4.38 (0.91)	4.14 (0.92)	3.79 (0.93)	3.33 (0.92)	2.99 (0.86)	2.62 (0.99)	2.26 (1.04)	2.01 (1.06)	1.60 (0.88)	1.32 (0.76)

Table 4.26: Descriptive statistics for valence across different densities, with mean values and standard deviations in parentheses. The table from the main study indicates that mean values generally decrease with density for 2D, 3D, and particle stimuli, with standard deviations reflecting varying degrees of variability.

Analysis of Valence Across Densities

Valence exhibits a progressive reduction in pleasant sentiments as density increases. Overall, as density increases for all avatar types, the perception of valence diminishes, although the severity varies. The standard deviation remains relatively constant, suggesting that the level of variability in responses is similar across various densities and avatar types.

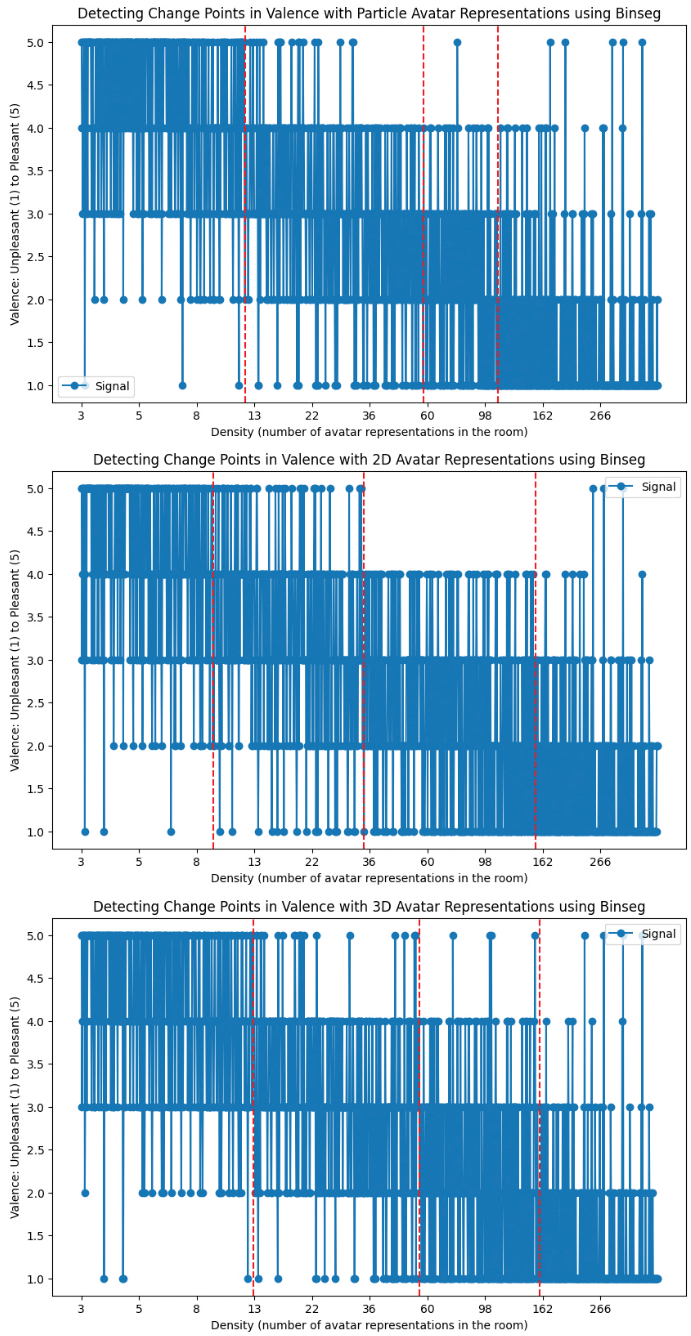

Figure 4.17: Detecting Change Points in Valence with Particle, 2D and 3D Avatar Representations using Binseg. The figure illustrates the detection of change points in valence with avatar representations using the Binary Segmentation (Binseg) method. The red dashed lines represent the identified change points.

The descriptive statistics for valence across different densities and stimuli types are summarized in Table 4.16. For the 2D stimuli, the mean valence values ranged from 4.38 ($SD = 0.90$) at a density of 3 to 1.43 ($SD = 0.78$) at a density of 266. For the 3D stimuli, the mean values ranged from 4.38 ($SD = 0.91$) at a density of 3 to 1.32 ($SD = 0.76$) at a density of 266. Lastly, for the particle stimuli, the mean values ranged from 4.36 ($SD = 0.84$) at a density of 3 to 1.30 ($SD = 0.81$) at a density of 266.

The Figure 4.16 illustrates the detection of change points in valence with particle, 2D and 3D avatar representations using the Binary Segmentation (Binseg) method. The x-axis represents the density, while the y-axis represents the valence scores, ranging from unpleasant to pleasant (1-5). The blue points indicate the signal, with red dashed lines marking the identified change points.

For particle avatar representations, change points were detected at the densities 8, 36, and 98. The data shows that between densities 3 and 8, participants reported feelings ranging from "Pleasant" to "Neutral," indicating higher levels of valence. Between densities 13 and 60, participants reported feelings from "Pleased" to "Neutral." Between densities 60 and 98, participants generally reported the valence level as "Neutral" to "Unsatisfied." Above density 98, there was a noticeable shift, with many participants rating the environment as "Unsatisfied" to "Unpleasant."

For 2D avatar representations, change points were detected at the densities 8, 22, and 98. The data shows that between densities 3 and 8, participants reported feelings ranging from "Pleasant" to "Neutral," indicating higher levels of valence. Between densities 8 and 36, participants reported high feelings from "Pleased" to "Neutral." Between densities 36 and 98, participants reported the valence level as "Neutral" to "Unsatisfied." Above density 98, there was a noticeable shift, with many participants rating the environment as "Unsatisfied" to "Unpleasant."

For 3D avatar representations, change points were detected at the densities 8, 36, and 98. The data shows that between densities 3 and 8, participants reported feelings ranging from "Pleasant" to "Neutral," indicating higher levels of valence. Between densities 8 and 36, participants reported high feelings from "Pleased" to "Neutral." Between densities 36 and 98, participants reported the valence level as "Neutral" to "Unsatisfied." Above density 98, there was a noticeable shift, with many participants rating the environment as "Unsatisfied" to "Unpleasant."

Assessing differences between avatar representation

To find out if there are differences between different types of avatar representations, we use the Friedman Test. As mentioned in Chapter 3. If the Friedman Test tells us that the rankings are different, we then use another test called the Wilcoxon signed-rank test as post-hoc to see exactly which avatars are different from each other.

density	χ^2	*p_value*	mean particle (rank)	mean 2D (rank)	mean 3D (rank)
3	2.571	0.276	1.24 (1st)	1.21 (2nd)	1.21 (2nd)
5	0.292	0.864	1.58 (2nd)	1.60 (1st)	1.56 (3rd)
8	14.652	0.001**	2.27 (1st)	2.06 (2nd)	2.06 (3rd)
13	2.687	0.261	2.80 (1st)	2.70 (3rd)	2.72 (2nd)
22	4.583	0.101	3.36 (1st)	3.22 (3rd)	3.24 (2nd)
36	0.933	0.627	3.65 (3rd)	3.69 (1st)	3.67 (2nd)
60	2.045	0.360	4.18 (1st)	4.15 (2nd)	4.07 (3rd)
98	30.618	0.001**	4.71 (1st)	4.44 (3rd)	4.45 (2nd)
162	25.025	0.001**	4.90 (1st)	4.72 (3rd)	4.77 (2nd)
266	10.778	0.005**	4.95 (1st)	4.86 (3rd)	4.93 (2nd)

Table 4.27: Friedman Test and Mean Ranks for Crowding Perception. The table presents the results of the main study, showing the chi-square (χ^2) values, p-values (*p_value*), and mean ranks for particle, 2D, and 3D stimuli across different densities.

The Table 4.27 includes columns for density, chi-square value (χ^2), p-value, and the mean ranks for each type of stimuli (particle, 2D, and 3D) for crowding perception. Significant p-values (less than 0.05) are marked with an asterisk (*) or double asterisk (**) to indicate their level of significance. Figure 4.18 illustrate the mean raking through all the densities for crowding perception.

For density 8 and the crowding perception, the non-parametric Friedman test of differences among repeated measures rendered a chi-square value of 14.652, which was significant ($p < 0.001$). To investigate the significant differences identified by the Friedman test, post-hoc pairwise comparisons were conducted using the Wilcoxon signed-rank test with Holm correction for multiple comparisons. The following pairs of avatar representations were compared: 2D vs. 3D, 2D vs. Particle, and 3D vs. Particle.

The Wilcoxon signed-rank test results are as follows: For the comparison between 2D and 3D avatars, $W = 818.0$, unadjusted $p = 0.941$, Holm-adjusted $p = 0.941$. For the comparison between 2D and Particle avatars, $W = 658.0$, unadjusted $p = 0.000877$, Holm-adjusted $p = 0.00263$. For the comparison between 3D and Particle avatars, $W = 609.0$, unadjusted $p = 0.000916$, Holm-adjusted $p = 0.00263$. After

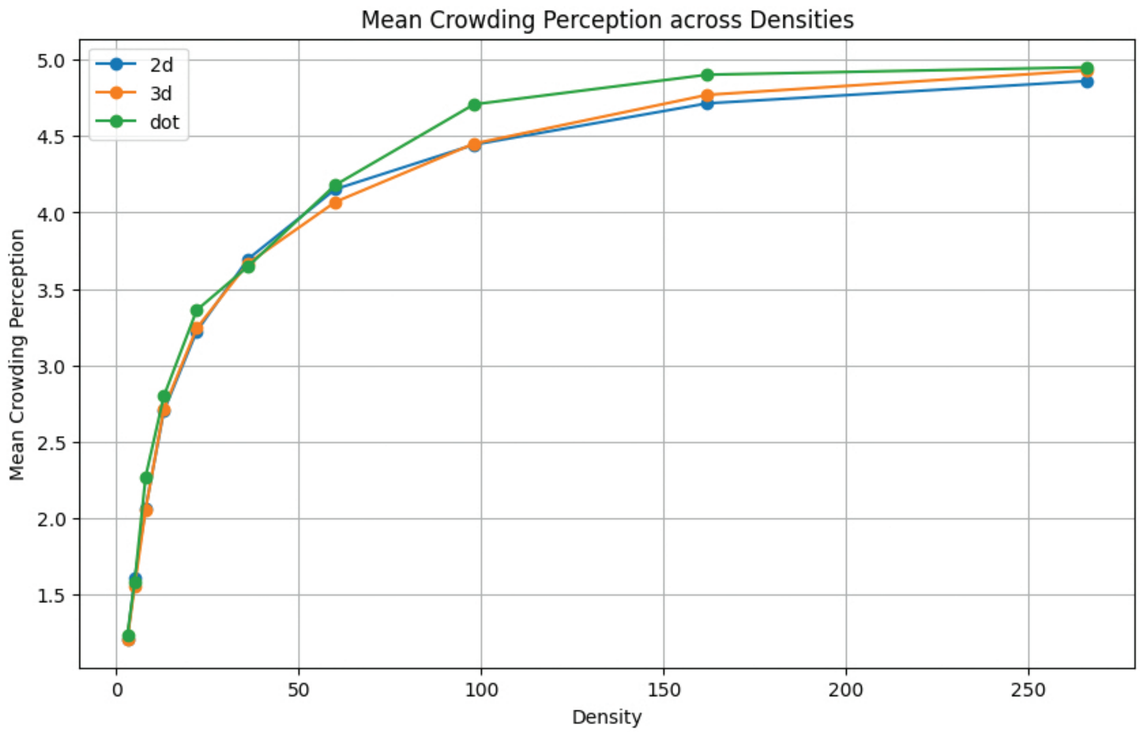

Figure 4.18: Mean crowding perception across densities. The image illustrates the relationship between density (number of avatar representations in the room) and mean crowding perception for three different stimuli types: 2D, 3D, and particle (dot) avatars. The x-axis represents the density, ranging from 0 to 266, while the y-axis shows the mean crowding perception, ranging from 1 to 5.

applying the Holm correction, the differences between the 2D and Particle avatars and between the 3D and Particle avatars were statistically significant at the $\alpha = 0.05$ level. The difference between the 2D and 3D avatars was not statistically significant.

For density 98 and the crowding perception, the non-parametric Friedman test of differences among repeated measures rendered a chi-square value of 30.618, which was significant ($p < 0.001$). To investigate the significant differences identified by the Friedman test, post-hoc pairwise comparisons were conducted using the Wilcoxon signed-rank test with Holm correction for multiple comparisons. The following pairs of avatar representations were compared: 2D vs. 3D, 2D vs. Particle, and 3D vs. Particle.

The Wilcoxon signed-rank test results are as follows: For the comparison between 2D and Particle avatars, $W = 176.0$, unadjusted $p = 0.000003$, Holm-adjusted $p = 0.000010$. For the comparison between 3D and Particle avatars, $W = 220.5$, unadjusted $p = 0.000003$, Holm-adjusted $p = 0.000010$. For the comparison between 2D and 3D avatars, $W = 421.0$, unadjusted $p = 0.892053$, Holm-adjusted $p = 0.892053$. After applying the Holm correction, the differences between the 2D and Particle avatars, and 3D and Particle avatars were statistically significant at the $\alpha = 0.05$ level. The difference between the 2D and 3D avatars was not statistically significant.

For density 162 and the crowding perception, the non-parametric Friedman test of differences among repeated measures rendered a chi-square value of 25.025, which was significant ($p < 0.001$). To investigate the significant differences identified by the Friedman test, post-hoc pairwise comparisons were conducted using the Wilcoxon signed-rank test with Holm correction for multiple comparisons. The following pairs of avatar representations were compared: 2D vs. 3D, 2D vs. Particle, and 3D vs. Particle.

The Wilcoxon signed-rank test results are as follows: For the comparison between 2D and Particle avatars, $W = 0.0$, unadjusted $p = 0.000001$, Holm-adjusted $p = 0.000004$. For the comparison between 3D and Particle avatars, $W = 44.0$, unadjusted $p = 0.001613$, Holm-adjusted $p = 0.003226$. For the comparison between 2D and 3D avatars, $W = 187.0$, unadjusted $p = 0.108226$, Holm-adjusted $p = 0.108226$. After applying the Holm correction, the differences between the 2D and Particle avatars, and 3D and Particle avatars were statistically significant at the $\alpha = 0.05$ level. The difference between the 2D and 3D avatars was not statistically significant.

For density 266 and the crowding perception, the non-parametric Friedman test of differences among repeated measures rendered a chi-square value of 10.778, which was significant ($p = 0.005$). To investigate the significant differences identified by the Friedman test, post-hoc pairwise comparisons were conducted using the Wilcoxon signed-rank test with Holm correction for multiple comparisons. The following pairs of avatar representations were compared: 2D vs. 3D, 2D vs. Particle, and 3D vs. Particle.

The Wilcoxon signed-rank test results are as follows: For the comparison between 2D and Particle avatars, $W = 24.0$, unadjusted $p = 0.006789$, Holm-adjusted $p = 0.020366$. For the comparison between 3D and Particle avatars, $W = 3.0$, unadjusted $p = 0.179712$, Holm-adjusted $p = 0.179712$. For the comparison between 2D and 3D avatars, $W = 19.5$, unadjusted $p = 0.025373$, Holm-adjusted $p = 0.050745$. After applying the Holm correction, the differences between the 2D and Particle avatars were statistically significant at the $\alpha = 0.05$ level. The differences between the 3D and Particle avatars, and 2D and 3D avatars were not statistically significant.

density	χ^2	p_value	mean particle (rank)	mean 2D (rank)	mean 3D (rank)
3	2.121	0.346	1.76 (3rd)	1.79 (2nd)	1.86 (1st)
5	6.379	0.041*	1.94 (1st)	1.85 (3rd)	1.86 (2nd)
8	2.267	0.322	2.10 (1st)	2.06 (2nd)	2.02 (3rd)
13	1.964	0.375	2.42 (1st)	2.31 (3rd)	2.37 (2nd)
22	6.536	0.038*	2.80 (1st)	2.62 (3rd)	2.71 (2nd)
36	11.751	0.003**	2.93 (2nd)	2.83 (3rd)	3.10 (1st)
60	8.489	0.014*	3.35 (1st)	3.15 (3rd)	3.28 (2nd)
98	15.696	< 0.001**	3.73 (1st)	3.51 (3rd)	3.56 (2nd)
162	15.296	< 0.001**	3.85 (2nd)	3.64 (3rd)	3.87 (1st)
266	25.532	< 0.001**	4.07 (2nd)	3.83 (3rd)	4.11 (1st)

Table 4.28: Friedman Test and Mean Ranks for Arousal. The table presents the results of the main study, showing the chi-square (χ^2) values, p-values (p_value), and mean ranks for particle, 2D, and 3D stimuli across different densities.

The Table 4.28 includes columns for density, chi-square value (χ^2), p-value, and the mean ranks for each type of stimuli (particle, 2D, and 3D) for arousal. Significant p-values (less than 0.05) are marked with an asterisk (*) or double asterisk (**) to indicate their level of significance. Figure 4.19 illustrate the mean raking through all the densities for arousal scores.

For density 5 and the arousal attribute, the non-parametric Friedman test of differences among repeated measures rendered a chi-square value of 6.379, which was significant ($p = 0.041$). To investigate the significant differences identified by the Friedman test, post-hoc pairwise comparisons were conducted using the Wilcoxon signed-rank test with Holm correction for multiple comparisons. The following pairs of avatar representations were compared: 2D vs. 3D, 2D vs. Particle, and 3D vs. Particle.

The Wilcoxon signed-rank test results are as follows: For the comparison between 2D and Particle avatars, $W = 363.0$, unadjusted $p = 0.161362$, Holm-adjusted $p = 0.484085$. For the comparison between 3D and Particle avatars, $W = 562.0$, unadjusted $p = 0.207586$, Holm-adjusted $p = 0.484085$. For the comparison between 2D and 3D avatars, $W = 474.0$, unadjusted $p = 0.599827$, Holm-adjusted $p = 0.599827$. After applying the Holm correction, none of the differences were statistically significant at the $\alpha = 0.05$ level.

For density 22 and the arousal attribute, the non-parametric Friedman test of differences among repeated measures rendered a chi-square value of 6.536, which was significant ($p = 0.038$). To investigate the significant differences identified by the Friedman test, post-hoc pairwise comparisons were conducted using the Wilcoxon signed-rank test with Holm correction for multiple comparisons. The following pairs

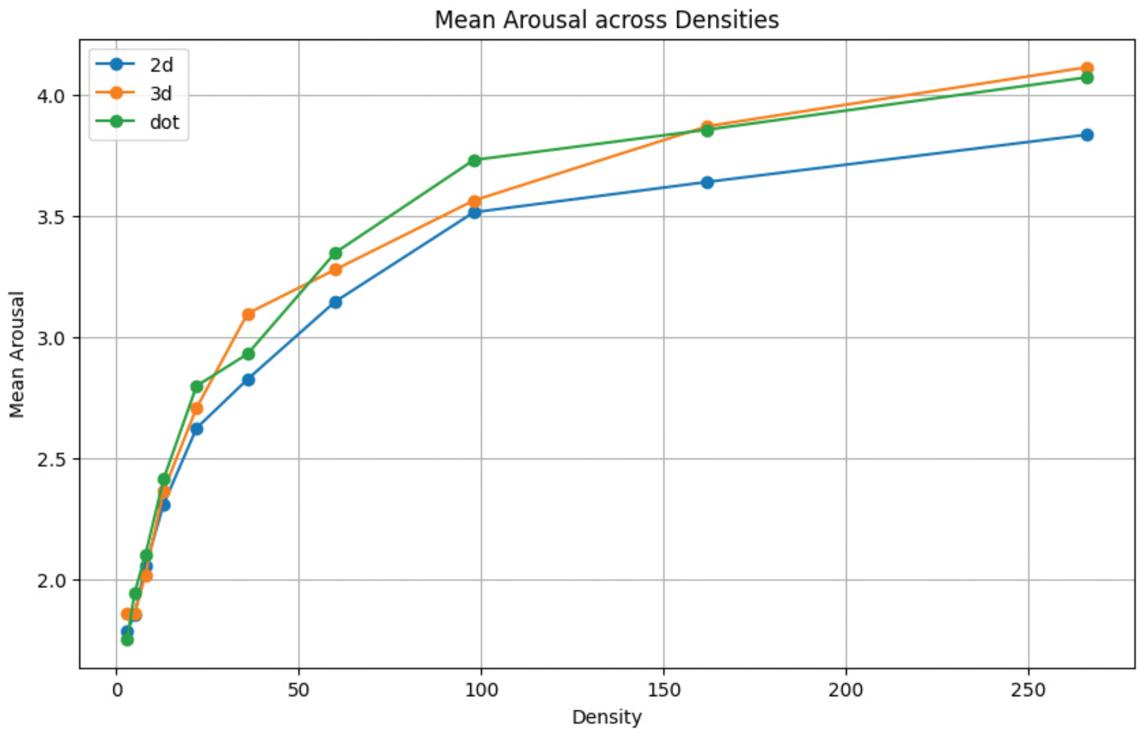

Figure 4.19: Mean arousal score across densities. The image illustrates the relationship between density (number of avatar representations in the room) and mean arousal scores for three different stimuli types: 2D, 3D, and particle (dot) avatars. The x-axis represents the density, ranging from 0 to 266, while the y-axis shows the mean arousal, ranging from 1 to 5.

of avatar representations were compared: 2D vs. 3D, 2D vs. Particle, and 3D vs. Particle.

The Wilcoxon signed-rank test results are as follows: For the comparison between 2D and Particle avatars, $W = 905.5$, unadjusted $p = 0.023160$, Holm-adjusted $p = 0.069479$. For the comparison between 3D and Particle avatars, $W = 935.5$, unadjusted $p = 0.169573$, Holm-adjusted $p = 0.339145$. For the comparison between 2D and 3D avatars, $W = 903.5$, unadjusted $p = 0.229077$, Holm-adjusted $p = 0.339145$. After applying the Holm correction, none of the differences were statistically significant at the $\alpha = 0.05$ level.

For density 36 and the arousal attribute, the non-parametric Friedman test of differences among repeated measures rendered a chi-square value of 11.751, which was significant ($p = 0.003$). To investigate the significant differences identified by the Friedman test, post-hoc pairwise comparisons were conducted using the Wilcoxon signed-rank test with Holm correction for multiple comparisons. The following pairs of avatar representations were compared: 2D vs. 3D, 2D vs. Particle, and 3D vs. Particle.

The Wilcoxon signed-rank test results are as follows: For the comparison between 2D and Particle avatars, $W = 1056.0$, unadjusted $p = 0.158094$, Holm-adjusted $p = 0.158094$. For the comparison between 3D and Particle avatars, $W = 805.5$, unadjusted $p = 0.014719$, Holm-adjusted $p = 0.029438$. For the comparison between 2D and 3D avatars, $W = 323.0$, unadjusted $p = 0.000224$, Holm-adjusted $p = 0.000671$. After applying the Holm correction, the differences between the 3D and Particle avatars and between the 2D and 3D avatars were statistically significant at the $\alpha = 0.05$ level. The difference between the 2D and Particle avatars was not statistically significant.

For density 60 and the arousal attribute, the non-parametric Friedman test of differences among repeated measures rendered a chi-square value of 8.489, which was significant ($p = 0.014$). To investigate the significant differences identified by the Friedman test, post-hoc pairwise comparisons were conducted using the Wilcoxon signed-rank test with Holm correction for multiple comparisons. The following pairs of avatar representations were compared: 2D vs. 3D, 2D vs. Particle, and 3D vs. Particle.

The Wilcoxon signed-rank test results are as follows: For the comparison between 2D and Particle avatars, $W = 670.5$, unadjusted $p = 0.007457$, Holm-adjusted $p = 0.022372$. For the comparison between 3D and Particle avatars, $W = 881.0$, unadjusted $p = 0.255911$, Holm-adjusted $p = 0.291581$. For the comparison between 2D and 3D avatars, $W = 860.0$, unadjusted $p = 0.145790$, Holm-adjusted $p = 0.291581$. After applying the Holm correction, the difference between the 2D and Particle avatars was statistically significant at the $\alpha = 0.05$ level. The differences between the 3D and Particle avatars and between the 2D and 3D avatars were not statistically significant.

For density 98 and the arousal attribute, the non-parametric Friedman test of differences among repeated measures rendered a chi-square value of 15.696, which was significant ($p < 0.001$). To investigate the significant differences identified by the Friedman test, post-hoc pairwise comparisons were conducted using the Wilcoxon signed-rank test with Holm correction for multiple comparisons. The following pairs of avatar representations were compared: 2D vs. 3D, 2D vs. Particle, and 3D vs. Particle.

The Wilcoxon signed-rank test results are as follows: For the comparison between 2D and Particle avatars, $W = 600.5$, unadjusted $p = 0.001757$, Holm-adjusted $p = 0.005272$. For the comparison between 3D and Particle avatars, $W = 734.0$, unadjusted $p = 0.029332$, Holm-adjusted $p = 0.058664$. For the comparison between 2D and 3D avatars, $W = 589.0$, unadjusted $p = 0.331677$, Holm-adjusted $p = 0.331677$. After applying the Holm correction, the difference between the 2D and Particle avatars was statistically significant at the $\alpha = 0.05$ level. The differences between the 3D and Particle avatars and between the 2D and 3D avatars were not statistically significant.

For density 162 and the arousal attribute, the non-parametric Friedman test of differences among repeated measures rendered a chi-square value of 15.296, which was significant ($p < 0.001$). To investigate the significant differences identified by the Friedman test, post-hoc pairwise comparisons were conducted using the Wilcoxon signed-rank test with Holm correction for multiple comparisons. The following pairs of avatar representations were compared: 2D vs. 3D, 2D vs. Particle, and 3D vs. Particle.

The Wilcoxon signed-rank test results are as follows: For the comparison between 2D and Particle avatars, $W = 535.5$, unadjusted $p = 0.005344$, Holm-adjusted $p = 0.010688$. For the comparison between 3D and Particle avatars, $W = 712.0$, unadjusted $p = 0.974449$, Holm-adjusted $p = 0.974449$. For the comparison between 2D and 3D avatars, $W = 279.0$, unadjusted $p = 0.002735$, Holm-adjusted $p = 0.008205$. After applying the Holm correction, the differences between the 2D and Particle avatars and between the 2D and 3D avatars were statistically significant at the $\alpha = 0.05$ level. The difference between the 3D and Particle avatars was not statistically significant.

For density 266 and the arousal attribute, the non-parametric Friedman test of differences among repeated measures rendered a chi-square value of 25.532, which was significant ($p < 0.001$). To investigate the significant differences identified by the Friedman test, post-hoc pairwise comparisons were conducted using the Wilcoxon signed-rank test with Holm correction for multiple comparisons. The following pairs of avatar representations were compared: 2D vs. 3D, 2D vs. Particle, and 3D vs. Particle.

The Wilcoxon signed-rank test results are as follows: For the comparison between 2D and Particle avatars, $W = 407.5$, unadjusted $p = 0.004667$, Holm-adjusted $p = 0.009335$. For the comparison between 3D and Particle avatars, $W = 313.5$, unadjusted $p = 0.747912$, Holm-adjusted $p = 0.747912$. For the comparison between 2D and 3D avatars, $W = 301.5$, unadjusted $p = 0.000668$, Holm-adjusted $p = 0.002003$. After applying the Holm correction, the differences between the 2D and Particle avatars and between the 2D and 3D avatars were statistically significant at the $\alpha = 0.05$ level. The difference between the 3D and Particle avatars was not statistically significant.

density	χ^2	*p_value*	**mean particle (rank)**	**mean 2D (rank)**	**mean 3D (rank)**
3	1.823	0.402	3.39 (2nd)	3.44 (1st)	3.32 (3rd)
5	1.638	0.441	3.40 (1st)	3.35 (2nd)	3.29 (3rd)
8	3.255	0.196	3.19 (2nd)	3.40 (1st)	3.19 (2nd)
13	2.474	0.290	2.99 (2nd)	3.10 (1st)	2.96 (3rd)
22	3.135	0.209	2.94 (1st)	2.90 (2nd)	2.79 (3rd)
36	6.969	0.031*	2.73 (1st)	2.73 (1st)	2.58 (3rd)
60	3.211	0.201	2.58 (2nd)	2.72 (1st)	2.58 (2nd)
98	0.858	0.651	2.43 (2nd)	2.47 (1st)	2.41 (3rd)
162	0.528	0.768	2.31 (2nd)	2.29 (3rd)	2.38 (1st)
266	1.511	0.470	2.35 (1st)	2.34 (2nd)	2.24 (3rd)

Table 4.29: Friedman Test and Mean Ranks for Dominance. The table presents the results of the main study, showing the chi-square (χ^2) values, p-values (*p_value*), and mean ranks for particle, 2D, and 3D stimuli across different densities.

The Table 4.29 includes columns for density, chi-square value (χ^2), p-value, and the mean ranks for each type of stimuli (particle, 2D, and 3D) for dominance. Significant p-values (less than 0.05) are marked with an asterisk (*) or double asterisk (**) to indicate their level of significance. Figure 4.20 illustrate the mean raking through all the densities for dominance scores.

For density 36 and the dominance attribute, the non-parametric Friedman test of differences among repeated measures rendered a chi-square value of 6.969, which was significant ($p = 0.031$). To investigate the significant differences identified by the Friedman test, post-hoc pairwise comparisons were conducted using the Wilcoxon signed-rank test with Holm correction for multiple comparisons. The following pairs of avatar representations were compared: 2D vs. 3D, 2D vs. Particle, and 3D vs. Particle.

The Wilcoxon signed-rank test results are as follows: For the comparison between 2D and Particle avatars, $W = 898.0$, unadjusted $p = 0.892549$, Holm-adjusted $p = 0.892549$. For the comparison between 3D and Particle avatars, $W = 622.0$, unadjusted $p = 0.020506$, Holm-adjusted $p = 0.061518$. For the comparison between 2D and 3D avatars, $W = 554.0$, unadjusted $p = 0.029573$, Holm-adjusted $p = 0.061518$.

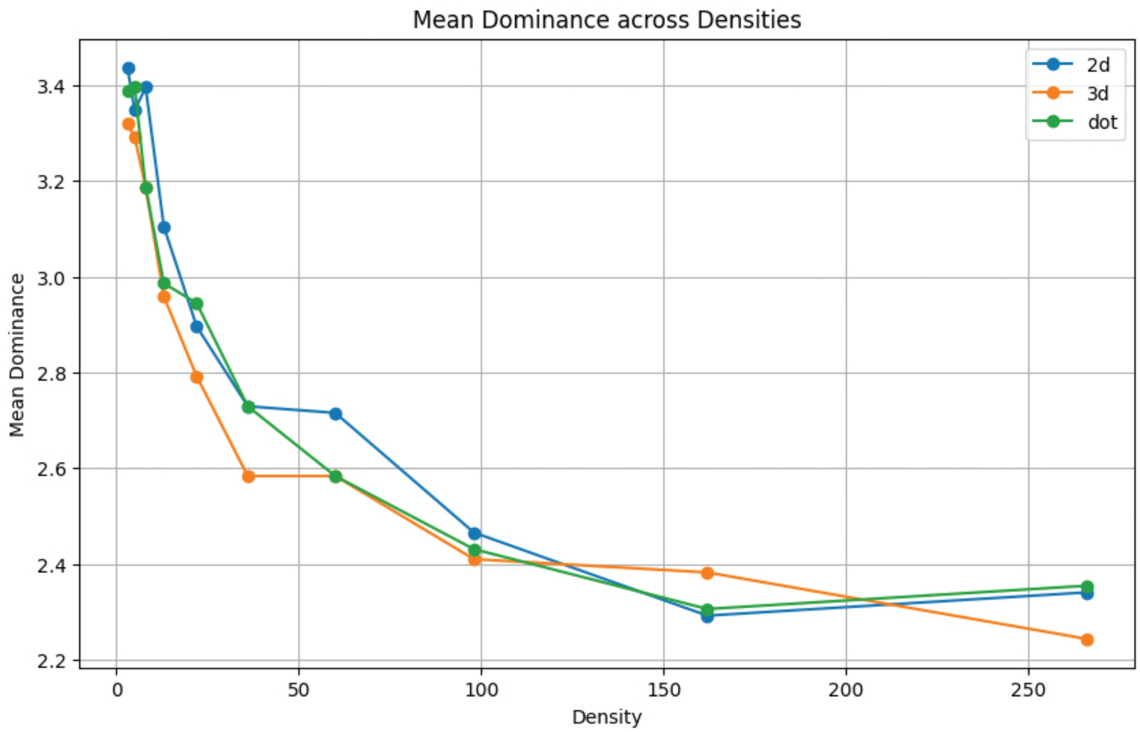

Figure 4.20: Mean dominance score across densities. The image illustrates the relationship between density (number of avatar representations in the room) and mean dominance scores for three different stimuli types: 2D, 3D, and particle (dot) avatars. The x-axis represents the density, ranging from 0 to 266, while the y-axis shows the mean dominance, ranging from 1 to 5.

After applying the Holm correction, the differences between the 2D and Particle avatars, 3D and Particle avatars, and 2D and 3D avatars were not statistically significant at the $\alpha = 0.05$ level.

The Table 4.30 includes columns for density, chi-square value (χ^2), p-value, and the mean ranks for each type of stimuli (particle, 2D, and 3D) for valence. Significant p-values (less than 0.05) are marked with an asterisk (*) or double asterisk (**) to indicate their level of significance. Figure 4.21 illustrate the mean raking through all the densities for valence scores.

Density	χ^2	*p_value*	Mean Particle (rank)	Mean 2D (rank)	Mean 3D (rank)
3	0.363	0.834	4.36 (3rd)	4.38 (1st)	4.38 (1st)
5	5.855	0.054	4.21 (1st)	4.01 (3rd)	4.14 (2nd)
8	0.518	0.772	3.74 (3rd)	3.76 (2nd)	3.79 (1st)
13	1.559	0.459	3.26 (2nd)	3.26 (2nd)	3.33 (1st)
22	5.946	0.051	2.88 (3rd)	2.96 (2nd)	2.99 (1st)
36	0.462	0.794	2.57 (3rd)	2.61 (2nd)	2.62 (1st)
60	0.574	0.751	2.24 (3rd)	2.31 (1st)	2.26 (2nd)
98	24.675	0.001**	1.75 (3rd)	2.02 (1st)	2.01 (2nd)
162	15.647	0.001**	1.49 (3rd)	1.72 (1st)	1.60 (2nd)
266	13.857	0.001**	1.30 (3rd)	1.43 (1st)	1.32 (2nd)

Table 4.30: Friedman Test and Mean Ranks for Valence. The table presents the results of the main study, showing the chi-square (χ^2) values, p-values (*p_value*), and mean ranks for particle, 2D, and 3D stimuli across different densities.

For density 5 and the valence attribute, the non-parametric Friedman test of differences among repeated measures rendered a chi-square value of 5.855, which was not significant ($p = 0.054$). To investigate the significant differences identified by the Friedman test, post-hoc pairwise comparisons were conducted using the Wilcoxon signed-rank test with Holm correction for multiple comparisons. The following pairs of avatar representations were compared: 2D vs. 3D, 2D vs. Particle, and 3D vs. Particle.

The Wilcoxon signed-rank test results are as follows: For the comparison between 2D and Particle avatars, $W = 421.0$, unadjusted $p = 0.005410$, Holm-adjusted $p = 0.016231$. For the comparison between 3D and Particle avatars, $W = 660.0$, unadjusted $p = 0.326607$, Holm-adjusted $p = 0.326607$. For the comparison between 2D and 3D avatars, $W = 516.5$, unadjusted $p = 0.148993$, Holm-adjusted $p = 0.297987$. After applying the Holm correction, the difference between the 2D and Particle avatars was statistically significant at the $\alpha = 0.05$ level. The differences between the 3D and Particle avatars and between the 2D and 3D avatars were not statistically significant.

For density 98 and the valence attribute, the non-parametric Friedman test of differences among repeated measures rendered a chi-square value of 24.675, which was significant ($p < 0.001$). To investigate the significant differences identified by the Friedman test, post-hoc pairwise comparisons were conducted using the Wilcoxon signed-rank test with Holm correction for multiple comparisons. The following pairs of avatar representations were compared: 2D vs. 3D, 2D vs. Particle, and 3D vs. Particle.

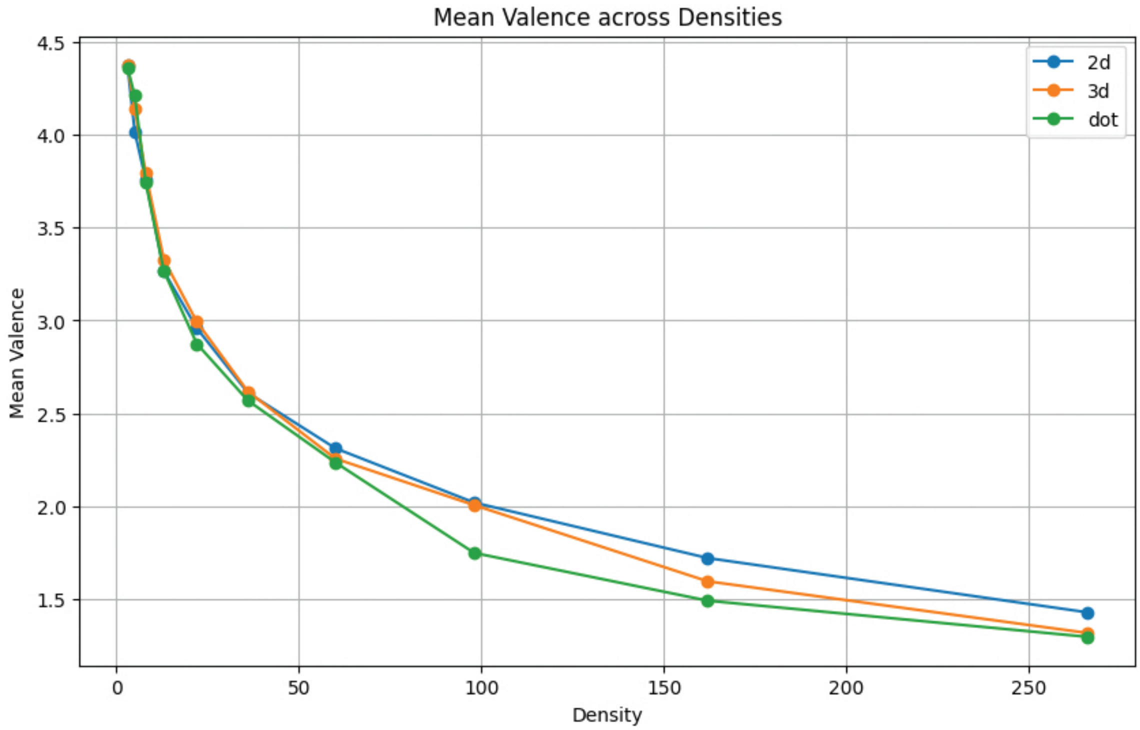

Figure 4.21: Mean valence score across densities. The image illustrates the relationship between density (number of avatar representations in the room) and mean valence scores for three different stimuli types: 2D, 3D, and particle (dot) avatars. The x-axis represents the density, ranging from 0 to 266, while the y-axis shows the mean valence, ranging from 1 to 5.

The Wilcoxon signed-rank test results are as follows: For the comparison between 2D and Particle avatars, $W = 289.5$, unadjusted $p = 0.000052$, Holm-adjusted $p = 0.000157$. For the comparison between 3D and Particle avatars, $W = 545.5$, unadjusted $p = 0.000686$, Holm-adjusted $p = 0.001372$. For the comparison between 2D and 3D avatars, $W = 466.0$, unadjusted $p = 0.717968$, Holm-adjusted $p = 0.717968$. After applying the Holm correction, the differences between the 2D and Particle avatars and between the 3D and Particle avatars were statistically significant at the $\alpha = 0.05$ level. The difference between the 2D and 3D avatars was not statistically significant.

For density 162 and the valence attribute, the non-parametric Friedman test of differences among repeated measures rendered a chi-square value of 15.647, which was significant ($p < 0.001$). To investigate the significant differences identified by the Friedman test, post-hoc pairwise comparisons were conducted using the Wilcoxon signed-rank test with Holm correction for multiple comparisons. The following pairs of avatar representations were compared: 2D vs. 3D, 2D vs. Particle, and 3D vs. Particle.

The Wilcoxon signed-rank test results are as follows: For the comparison between 2D and Particle avatars, $W = 283.5$, unadjusted $p = 0.000240$, Holm-adjusted $p = 0.000719$. For the comparison between 3D and Particle avatars, $W = 484.5$, unadjusted $p = 0.073384$, Holm-adjusted $p = 0.146551$. For the comparison between 2D and 3D avatars, $W = 461.5$, unadjusted $p = 0.073276$, Holm-adjusted $p = 0.146551$. After applying the Holm correction, the difference between the 2D and Particle avatars was statistically significant at the $\alpha = 0.05$ level. The differences between the 3D and Particle avatars and between the 2D and 3D avatars were not statistically significant.

For density 266 and the valence attribute, the non-parametric Friedman test of differences among repeated measures rendered a chi-square value of 13.857, which was significant ($p = 0.001$). To investigate the significant differences identified by the Friedman test, post-hoc pairwise comparisons were conducted using the Wilcoxon signed-rank test with Holm correction for multiple comparisons. The following pairs of avatar representations were compared: 2D vs. 3D, 2D vs. Particle, and 3D vs. Particle.

The Wilcoxon signed-rank test results are as follows: For the comparison between 2D and Particle avatars, $W = 126.0$, unadjusted $p = 0.002308$, Holm-adjusted $p = 0.006923$. For the comparison between 3D and Particle avatars, $W = 94.5$, unadjusted $p = 0.414216$, Holm-adjusted $p = 0.414216$. For the comparison between 2D and 3D avatars, $W = 116.0$, unadjusted $p = 0.007633$, Holm-adjusted $p = 0.015267$. After applying the Holm correction, the differences between the 2D and Particle avatars and between the 2D and 3D avatars were statistically significant at the $\alpha = 0.05$ level. The difference between the 3D and Particle avatars was not statistically significant.

Exploring relationships

In Table 4.31, the intercorrelations between spatial crowding, arousal, valence, and dominance were showed. Spatial crowding was found to have a strong negative correlation with valence ($r = -0.72$), indicating that as perceived crowding increases, feelings of valence decrease. Conversely, spatial crowding had a positive correlation with arousal ($r = 0.24$), suggesting that higher levels of crowding are associated with increased arousal. The correlation between spatial crowding and dominance was weakly negative ($r = -0.05$), implying a negligible relationship. Arousal was positively and weakly correlated with dominance ($r = 0.14$) and weakly negatively correlated with valence ($r = -0.18$), indicating that higher arousal levels are slightly associated with lower valence and higher dominance. Valence showed a weak positive correlation with dominance ($r = 0.22$), suggesting that higher valence is moderately associated with higher dominance.

	1	2	3	
1. Spatial Crowding	1.00			
2. Arousal (SAM)	0.45	1.00		
3. Valence (SAM)	-0.79	-0.47	1.00	
4. Dominance (SAM)	-0.03	0.13	0.24	1.00

Table 4.31: Intercorrelations of Subjective Measures. Table presents the intercorrelations between subjective measures, including spatial crowding, arousal (SAM), valence (SAM), and dominance (SAM), showing the relationships between these variables.

Density	Attribute	2d	3d	Dot
3	Crowding Perception	0.780	0.803	0.866
3	Valence	0.909	0.846	0.808
3	Arousal	0.903	0.944	0.853
3	Dominance	0.877	0.867	0.886
5	Crowding Perception	0.918	0.872	0.812
5	Valence	0.810	0.908	0.863
5	Arousal	0.900	0.845	0.874
5	Dominance	0.858	0.817	0.862
8	Crowding Perception	0.881	0.825	0.796
8	Valence	0.894	0.889	0.762
8	Arousal	0.882	0.821	0.837
8	Dominance	0.923	0.868	0.900
13	Crowding Perception	0.859	0.833	0.849
13	Valence	0.815	0.762	0.796
13	Arousal	0.784	0.840	0.789
13	Dominance	0.812	0.795	0.868
22	Crowding Perception	0.927	0.868	0.844
22	Valence	0.883	0.850	0.822
22	Arousal	0.901	0.794	0.844
22	Dominance	0.899	0.817	0.864
36	Crowding Perception	0.789	0.892	0.758
36	Valence	0.814	0.903	0.798
36	Arousal	0.900	0.917	0.867
36	Dominance	0.853	0.879	0.812
60	Crowding Perception	0.815	0.807	0.764
60	Valence	0.881	0.817	0.872
60	Arousal	0.849	0.836	0.883
60	Dominance	0.895	0.811	0.874
98	Crowding Perception	0.900	0.804	0.902
98	Valence	0.874	0.876	0.873
98	Arousal	0.900	0.890	0.781
98	Dominance	0.909	0.867	0.853
162	Crowding Perception	0.717	0.616	0.758
162	Valence	0.928	0.757	0.773
162	Arousal	0.944	0.923	0.897
162	Dominance	0.855	0.883	0.931
266	Crowding Perception	0.674	0.826	0.898
266	Valence	0.874	0.932	0.889
266	Arousal	0.914	0.890	0.781
266	Dominance	0.909	0.872	0.853

Table 4.21: ICC(C, k) for Different Densities and Attributes. The table shows that ICC(C, k) values for different densities and attributes across 2D, 3D, and Particle stimuli types generally indicate high reliability, with most values above 0.7.

Chapter 5

Discussions

In this thesis, I examined how different avatar types and densities affect users' experiences in virtual spaces. Similar to physical environments, virtual crowding can influence users' comfort and emotional states, which is increasingly important as more interactions shift to digital platforms. This statement was not only mentioned in related work, but it was also confirmed by the results of my experiments. The purpose of this work was to gain a better understanding of how crowding affects experiences in these spaces and whether these effects align with traditional architectural principles.

My journey began with an interest in how digital spaces could mimic or diverge from the physical ones we are so accustomed to. To explore this, I adopted a within-subjects strategy, recruiting participants through various channels to ensure a diverse and representative sample for two experiments. Participants reviewed visual representations of rooms with varying degrees of crowding and avatar types, allowing me to gather robust and unbiased data. This methodological choice was crucial in capturing the nuances of how crowd density and avatar representation impact crowding perception.

As the data started to come in, intriguing patterns emerged. It became clear that users' perceptions of crowding and their emotional responses were not consistent across multiple exposures to varying crowd densities and different avatar representations. This inconsistency suggests that the type of avatar representation used might significantly influence these perceptions. Going deeper, I was able to identify thresholds and change points where emotional responses shifted significantly.

Moreover, different types of avatar representations—particle, 2D, and 3D—had distinct impacts on users' perceptions of crowding and their emotional responses. This finding was particularly compelling as it highlighted the complexity of virtual crowding dynamics and underscored the importance of thoughtful design in digital environments. These and other findings are discussed in greater detail in the following sections.

In summary, this thesis explores the impact of avatar density and representation in virtual environments on user perceptions and emotional reactions, thereby deepening our understanding of crowding perception. By meticulously studying various avatar types at different crowd densities, this research offers valuable insights into the emotional and psychological dynamics of virtual crowding. In the subsequent sections, I will connect these findings with the research questions posed in Chapter 3, discuss the limitations of the work, and demonstrate how the chosen methodology facilitated these new insights. Ultimately, I hope this work will provide a foundation for future research and offer practical recommendations for designing more user-friendly virtual environments.

—

These findings indicate that while there were generally consistent responses, certain densities and avatar representations exhibited significant inconsistencies.

Como designer, sabemos a importancia da consistencia para desenvolver projetos, mas ao mesmo tempo eh preciso ter consciencia que somos individuos unicos e que inconsistencia podem acontecer. Com esses findings, espero que facilite para entender melhor aonde exatamente essas inconsistencias acontecem, para facilitar que trabalhemos com ela.—

5.1 Connecting findings with research questions

This section presents the findings from the pilot and main studies, connecting the research questions, hypotheses, and results to provide an understanding of the impact of avatar density and representation on user perceptions and emotional reactions in virtual environments.

The pilot study ran from March 28, 2023, to October 12, 2023, with 70 participants. Using a within-subjects design, participants were shown each stimulus twice and were exposed to different experimental conditions that changed the density and type of avatar representation (particle, 2D, and 3D). The findings from the pilot study are summarized below. The main study ran from February 12, 2024, to March 16, 2024, with 48 participants. The study aimed to replicate and extend the pilot study findings, including more densities and showing each stimulus three times.

5.1.1 User Perception Consistency in Virtual Crowds: Avatar Type Matters

In our first research question, **RQ1: How consistent are users' crowding perceptions and emotional responses (valence, arousal, dominance) when exposed to varying crowd densities and different avatar representations (particle, 2D, and 3D) multiple times?**, my aim was to use the study design, with repeated images, as an opportunity to understand if there is consistency among participants with their own ratings (intrarater) and between participants (interrater). From this research question, several hypotheses emerged.

The null hypothesis (RQ1.H0) states that **if users are exposed to varying crowd densities with different avatar representations (particle, 2D, and 3D), their crowding perceptions and emotional responses (valence, arousal, and dominance) will remain consistent across multiple exposures.** The results showed that crowding perceptions and emotional responses did not remain consistent across multiple exposures for different avatar types and densities. For example, for 2D avatar representations, there were significant differences at density 5 ($p < .01$, $Z = 69$) and density 8 ($p = .03$, $Z = 235.5$) for crowding perception, and similar inconsistencies were observed for arousal and valence. Therefore, this hypothesis was rejected.

Going forward to the next hypothesis, (RQ1.H1) states that **if users are exposed to varying crowd densities with particle avatar representations, their crowding perceptions and emotional responses (valence, arousal, and dominance) will significantly differ across multiple exposures.** In the pilot study, the Wilcoxon signed-rank test results for particle avatar representations showed a significant difference in dominance at density 3 ($p = .04$, $Z = 42$). In the main study, the Friedman test showed statistically significant differences for particle avatars in valence at density 5 ($p = 0.002$), arousal at density 22 ($p = 0.006$), and crowding perception at density 98 ($p = 0.002$). This indicates that participants' emotional responses and crowding perceptions varied significantly across these densities when exposed to particle avatars, both in the pilot and main studies. These consistent findings across both studies support the hypothesis that user perceptions and emotional responses significantly differ with varying crowd densities and particle avatar representations.

Moving on to the next hypothesis, (RQ1.H2) states that **if users are exposed**

to varying crowd densities with 2D avatar representations, their crowding perceptions and emotional responses (valence, arousal, and dominance) will significantly differ across multiple exposures. In the pilot study, the Wilcoxon signed-rank test results for 2D avatar representations showed significant differences in several densities: crowding perception at density 5 ($p < .01$, $Z = 69$) and density 8 ($p = .03$, $Z = 235.5$), arousal at density 5 ($p = .02$, $Z = 92$) and density 8 ($p = .02$, $Z = 71$), dominance at density 22 ($p = .03$, $Z = 196$), and valence at density 3 ($p = .008$, $Z = 76$) and density 5 ($p = .02$, $Z = 75$). In the main study, the Friedman test showed statistically significant differences for 2D avatars in valence at density 8 ($p = 0.011$) and density 266 ($p = 0.025$), and crowding perception at density 60 ($p = 0.018$). These results indicate that participants' emotional responses and crowding perceptions varied significantly across these densities when exposed to 2D avatars, both in the pilot and main studies. These consistent findings across both studies support the hypothesis that user perceptions and emotional responses significantly differ with varying crowd densities and 2D avatar representations.

Finally, our last hypothesis for this research question, (RQ1.H3) states that **if users are exposed to varying crowd densities with 3D avatar representations, their crowding perceptions and emotional responses (valence, arousal, and dominance) will significantly differ across multiple exposures.** In the pilot study, the Wilcoxon signed-rank test results for 3D avatar representations revealed significant differences in dominance at density 22 ($p = .01$, $Z = 158$). In the main study, the Friedman test showed a statistically significant difference for 3D avatars in dominance at density 13 ($p = 0.006$). This indicates that participants' emotional responses and crowding perceptions varied significantly at these densities when exposed to 3D avatars, both in the pilot and main studies. It is interesting that while we found inconsistencies in ratings from 3D avatars in dominance, the most realistic avatar representation of a humanoid performed the best through different subjective measurements (crowding and self-assessment manikin). This suggests that using a more realistic avatar could improve the consistency of user experiences and perceptions.

The consistency of significant results across both studies demonstrates the replicability of the findings. Both the pilot and main studies reveal that user perceptions and emotional responses vary significantly with different avatar types and crowd densities, challenging the initial null hypothesis and underscoring the importance of avatar representation in virtual environments. These findings suggest that while there were generally consistent responses across repeated stimuli, inconsistencies were more pronounced at specific densities and for certain avatar representations, particularly 2D avatars.

5.1.2 Impact of Avatar Types on Crowd Density Perception

In our second research question, **RQ2: Is there a threshold density at which users' emotional responses significantly change?**, my aim was to explore the crowding perception literature, which suggests that crowding perception is affected when a threshold is reached. Density changes to crowding when it reaches a level that affects the behavior, physiology, or experience of individuals in an environment [54]. The author mentioned that this transition is determined by its consequences on individuals rather than a specific numerical threshold. My aim with this research question was to identify a specific numerical threshold in my density measurements. Binary Segmentation (Binseg) was used to detect change points in crowding perception, valence, arousal, and dominance for different avatar representations and for the ease of reading, the Table 5.1 was created comparing the studies.

The null hypothesis (RQ2.H0) states that **If avatars are represented in different formats, the thresholds for significant changes in users' emotional responses will remain the same regardless of the representation (particle, 2D, or 3D).** The pilot study results indicate that while there were some common densities (e.g., 8 and 162) where significant changes were detected, there were also unique thresholds for different avatars in certain cases. The main study results demonstrate that while some thresholds for significant changes in users' emotional responses were consistent across different avatar representations (notably at density 8), there were variations in specific thresholds for different attributes and avatar types. Therefore, the hypothesis that the thresholds for significant changes in users' emotional responses will remain the same regardless of the avatar representation is not fully supported. The findings suggest that while there are common thresholds, unique thresholds also exist for different avatar types, indicating that the format of avatar representation does influence the thresholds for changes in users' emotional responses.

Moving to the next hypothesis, **(RQ2.H1)** states that **If particle representations of avatars are used, the threshold density for significant changes in users' emotional responses will be higher compared to 2D avatar representations.** The findings from both the pilot and main studies provide valuable insights into this hypothesis.

In the pilot study, the change points for particle avatars were identified at densities 13 and 162 for crowding perception, 8 and 162 for arousal, 13 and 162 for dominance, and 8 and 162 for valence. For 2D avatars, the change points were at densities 8 and 162 for crowding perception, arousal, and dominance, and at densities 13 and 162 for valence. This indicates that particle avatars generally had higher thresholds for crowding perception and dominance compared to 2D avatars, partially supporting the hypothesis.

In the main study, which included a broader range of densities [3, 5, 8, 13, 22, 36, 60, 98, 162, 266], significant change points for particle avatars were found at densities 8, 13, and 60 for crowding perception; 8, 36, and 98 for valence; 8, 36, and 60 for arousal; and 8 and 60 for dominance. For 2D avatars, change points were at densities 8, 22, and 60 for crowding perception; 8, 22, and 98 for valence; 8, 60, and 162 for arousal; and 8 and 98 for dominance.

Comparing the thresholds with 2D avatars, particle avatars demonstrated higher thresholds for crowding perception (13 vs 8) and dominance (13 vs 8) in the pilot study. However, for valence, 2D avatars had a higher threshold (13 vs 8). In the main study, the first threshold was the same for both avatars across all measures (8). For the second threshold, 2D avatars had a higher threshold for crowding perception (22 vs 13) and arousal (60 vs 36), while particle avatars had higher thresholds for valence (36 vs 22). For dominance, 2D avatars had a higher second threshold (98 vs 60).

These results indicate that while particle avatars generally show higher thresholds for certain emotional responses compared to 2D avatars, the consistency is not absolute across all measures. The hypothesis is partially supported, particularly in the pilot study, but the main study reveals mixed results, suggesting that the relationship between avatar representation and emotional response thresholds is complex and influenced by multiple factors.

Avatar Type	Measure	Pilot Study	Main Study
Particle	Crowding	13, 162	8, 13, 60
	Valence	8, 162	8, 36, 98
	Arousal	8, 162	8, 36, 60
	Dominance	13, 162	8, 60
2D	Crowding	8, 162	8, 22, 60
	Valence	13, 162	8, 22, 98
	Arousal	8, 162	8, 60, 162
	Dominance	8, 162	8, 98
3D	Crowding	8, 162	8, 36, 98
	Valence	8, 162	8, 36, 98
	Arousal	8, 162	13, 36, 98
	Dominance	8, 162	13, 98

Table 5.1: Change Point Thresholds for Different Avatar Representations in Pilot and Main Studies

Moving to the next hypothesis, **(RQ2.H2)** states that **If particle representations of avatars are used, the threshold density for significant changes in users' emotional responses will be higher compared to 3D avatar representations.** The findings from both the pilot and main studies give insights into this hypothesis.

In the pilot study, the change points for 3D avatars were identified at densities 8 and 162 for crowding perception, arousal, and valence, and at densities 13 and 162 for dominance. Comparing these to particle avatars, the thresholds were generally similar, with some differences indicating that particle avatars had higher thresholds for crowding perception (13 vs 8) and dominance (13 vs 8). However, for arousal and valence, the thresholds were the same (8). Comparing the thresholds with 3D avatars, particle avatars generally showed higher thresholds for crowding perception and dominance in the pilot study. In the main study, 3D avatars had higher thresholds arousal and dominance (8 vs 13). These results indicate that the hypothesis is partially supported, particularly for crowding perception and dominance in the pilot study. However, the main study reveals that 3D had the same or higher thresholds across different emotional measures and the lower variability in 3D avatar responses suggests they might provide more consistent emotional cues across participants.

In our last hypothesis, **(RQ2.H3)** states that **If 2D avatars are used, the threshold density for significant changes in users' emotional responses will be higher compared to 3D avatar representations.** In the pilot study, the change points for 2D avatars were identified at densities 8 and 162 for crowding perception, arousal, and dominance, and at densities 13 and 162 for valence. For 3D avatars, the change points were at densities 8 and 162 for crowding perception, arousal, dominance, and valence. In the main study, significant change points for 2D avatars were found at densities 8, 22, and 60 for crowding perception, 8, 22, and 98 for valence, 8, 60, and 162 for arousal, and 8 and 98 for dominance. For 3D avatars, change points were at densities 8, 36, and 98 for crowding perception, 8, 36, and 98 for valence, 13, 36, and 98 for arousal, and 13 and 98 for dominance.

Comparing the thresholds with 3D avatars, 2D avatars demonstrated higher thresholds for valence (8 vs. 13) in the pilot study. In the main study, 3D avatars had higher thresholds arousal and dominance (8 vs 13) and lower when comparing arousal (98 vs 162). These results indicate that the hypothesis is partially supported.

At a high level, it was important to include more densities to reach a more granular threshold. With three thresholds and ten densities in the main study, it was possible to identify the middle point for neutral feelings more precisely. Additionally, density 8 emerged as a consistent threshold across all avatar types (particle, 2D, and 3D) for crowding perception, valence, and arousal. This suggests that at this specific density, users' perceptions and emotional responses undergo a notable shift regardless of the visual format of the avatars. The uniformity of this threshold indicates a fundamental aspect of how crowd density impacts user experience, highlighting its significance in the design of virtual environments. This consistency can be leveraged to predict user discomfort or changes in emotional state, allowing for proactive adjustments in crowd management and virtual space design.

Another interesting finding was the visualization of change points and inconsistencies in behavior. For example, crowding and valence showed a clearer distribution of change points, while dominance and arousal were more scattered and challenging to interpret. Moreover, 3D avatar representations exhibited less variation between change points for crowding, valence, arousal, and dominance. Understanding these common thresholds is crucial for designing virtual spaces that aim to maintain user comfort and positive experiences.

5.1.3 How Avatar Density and Representation Make the Difference

In our third research question, **RQ3: Do different types of avatar representations (particle, 2D, and 3D) and varying densities of avatars in a virtual**

environment lead to differences in users' perceptions of crowding and their emotional responses (valence, arousal, and dominance)?, my aim was to investigate whether the type of avatar representation and the density of avatars in a virtual environment significantly impact users' crowding perceptions and emotional responses. Previous studies have indicated that crowding perception can be influenced by various factors, including the type and density of visual stimuli in an environment. This research question seeks to identify specific differences across different avatar representations.

The null hypothesis (RQ3.H0) states that**If different types of avatar representations (particle, 2D, and 3D) are placed in an environment, there will be no significant differences in users' perceptions of crowding and their emotional responses (valence, arousal, and dominance).** In the pilot study, significant differences were found at various densities, indicating that the type of avatar representation does influence crowding perception. For instance, at density 8, the Friedman test revealed a chi-square value of 22.521 ($p < 0.001$), and post-hoc Wilcoxon tests showed significant differences between 2D and 3D avatars (Holm-adjusted $p = 0.000596$) and between 3D and particle avatars (Holm-adjusted $p = 0.000596$). These results suggest that 3D avatars lead to higher crowding perceptions compared to 2D and particle avatars, partially rejecting the null hypothesis.

In the main study, significant differences were also observed at densities 8, 98, and 162. For example, at density 8, the Friedman test rendered a chi-square value of 14.652 ($p < 0.001$), with significant differences found between 2D and particle avatars (Holm-adjusted $p = 0.00263$) and between 3D and particle avatars (Holm-adjusted $p = 0.00263$). These findings reinforce the notion that different avatar representations significantly impact crowding perception.

Moving to the next, the hypothesis (**RQ3.H1**) posits that **If particle representations of avatars are placed in an environment, users' perceptions of crowding and their emotional responses (valence, arousal, and dominance) will be less intense compared to when 2D avatars are used.** In the pilot study, significant differences in valence were identified at density 5, where the Friedman test rendered a chi-square value of 11.522 ($p = 0.003$), and post-hoc Wilcoxon tests showed significant differences between 2D and particle avatars (Holm-adjusted $p = 0.015256$). The main study revealed significant differences at densities 98 and 162. At density 98, the Friedman test rendered a chi-square value of 24.675 ($p < 0.001$), with significant differences found between 2D and particle avatars (Holm-adjusted $p = 0.000157$) and between 3D and particle avatars (Holm-adjusted $p = 0.001372$). Comparing these results, particle avatars generally led to higher valence scores (less negative emotional responses) compared to 2D avatars, partly supporting the hypothesis.

The hypothesis (**RQ3.H2**) states that **If particle representations of avatars are placed in an environment, users' perceptions of crowding and their emotional responses (valence, arousal, and dominance) will be less intense compared to when 3D avatars are used.** In the pilot study, significant differences in arousal were found at density 13, where the Friedman test rendered a chi-square value of 9.324 ($p = 0.009$), and post-hoc Wilcoxon tests showed significant differences between 2D and particle avatars (Holm-adjusted $p = 0.021385$) and between 2D and 3D avatars (Holm-adjusted $p = 0.021385$). In the main study, significant differences were found at densities 36 and 60. At density 36, the Friedman test rendered a chi-square value of 11.751 ($p = 0.003$), with significant differences found between 3D and particle avatars (Holm-adjusted $p = 0.029438$) and between 2D and 3D avatars (Holm-adjusted $p = 0.000671$). These results indicate that particle avatars often result in lower arousal levels compared to 3D avatars, partially supporting the hypothesis.

The hypothesis (**RQ3.H3**) states that **If 2D avatars are placed in an environment, users' perceptions of crowding and their emotional responses (valence, arousal, and dominance) will be less intense compared to when 3D avatars are used.** In the pilot study, significant differences in dominance were identified at density 5, where the Friedman test rendered a chi-square value of 7.224 ($p = 0.027$), but post-hoc Wilcoxon tests showed no significant differences between any pairs of avatars after Holm correction. In the main study, significant differences were found at density 36. The Friedman test rendered a chi-square value of 6.969 ($p = 0.031$), but post-hoc Wilcoxon tests showed no significant differences between any pairs of avatars after Holm correction. These results indicate that while there are some differences, the hypothesis is only partially supported, as 3D avatars did not consistently lead to more intense dominance responses compared to 2D avatars.

Taken together, our findings indicate different types of avatar representations and varying densities do lead to significant differences in users' perceptions of crowding and their emotional responses. The null hypothesis (**RQ3.H0**) is largely rejected, as significant differences were found across multiple densities and measures (crowding perception, valence, arousal, and dominance).

Particle avatars generally result in less intense emotional responses compared to 2D and 3D avatars, supporting hypotheses **RQ3.H1** and **RQ3.H2**. The support for **RQ3.H3** is partial, with 2D avatars sometimes leading to less intense responses compared to 3D avatars.

5.2 Linear Variations with Avatar Density

In our fourth research question, **RQ4: Is there a linear relationship between avatar density and perceived crowding, and how do emotions (valence, arousal, dominance) vary linearly with avatar density for different types of avatar representations?**, my aim was to investigate whether increasing avatar density in a virtual environment leads to a linear relationship with perceived crowding and how emotional responses (valence, arousal, and dominance) vary linearly with avatar density across different avatar representations (particle, 2D, and 3D).

The null hypothesis (**RQ4.H0**) states that **If avatar density increases, there will be no linear relationship between avatar density and perceived crowding, nor will there be a linear variation in emotional responses (valence, arousal, and dominance) for different types of avatar representations (particle, 2D, and 3D).** In the pilot study, significant linear relationships were found between avatar density and perceived crowding, as well as emotional responses, indicating that the type of avatar representation does influence these perceptions. For instance, spatial crowding had a strong negative correlation with valence ($r = -0.72$), and a positive correlation with arousal ($r = 0.24$). These results suggest that increasing avatar density leads to higher perceived crowding and varying emotional responses, partially rejecting the null hypothesis.

In the main study, significant linear relationships were also observed. For example, spatial crowding had a strong negative correlation with valence ($r = -0.79$) and a positive correlation with arousal ($r = 0.45$). These findings reinforce the notion that increasing avatar density significantly impacts perceived crowding and emotional responses.

Moving to the next hypothesis, **(RQ4.H1)** posits that **If avatar density increases, there will be a linear relationship where perceived crowding increases for particle representations of avatars.** In the pilot study, the results showed a significant positive linear relationship between avatar density and perceived crowding for particle avatars, supporting the hypothesis. The main study confirmed these findings, with a strong linear relationship observed between density and crowding perception for particle avatars.

The hypothesis (**RQ4.H2**) states that **If avatar density increases, there will be a linear relationship where perceived crowding increases for 2D representations of avatars.** Both the pilot and main studies showed significant positive linear relationships between avatar density and perceived crowding for 2D avatars, supporting the hypothesis. The linear trend was consistent across different densities, indicating that users' perceptions of crowding increased linearly with the density of 2D avatars.

The hypothesis (**RQ4.H3**) states that **If avatar density increases, there will be a linear relationship where perceived crowding increases for 3D representations of avatars.** The analysis for 3D avatars demonstrated significant positive linear relationships between avatar density and perceived crowding in both the pilot and main studies, supporting the hypothesis. The results confirmed that higher densities of 3D avatars lead to increased perceptions of crowding in a linear manner.

Moving to the emotional responses, the hypothesis (**RQ4.H4**) posits that **If avatar density increases, the emotional responses (valence, arousal, and dominance) will vary linearly for particle representations of avatars.** In the pilot study, a significant negative linear relationship was found between avatar density and valence for particle avatars, indicating that higher densities led to less positive emotional responses, supporting the hypothesis. Similar results were found for arousal, with a positive linear relationship observed.

The hypothesis (**RQ4.H5**) states that **If avatar density increases, the emotional responses (valence, arousal, and dominance) will vary linearly for 2D representations of avatars.** The main study showed that increasing density led to a significant linear decrease in valence and a linear increase in arousal for 2D avatars, supporting the hypothesis. These findings were consistent with the pilot study results.

The hypothesis (**RQ4.H6**) posits that **If avatar density increases, the emotional responses (valence, arousal, and dominance) will vary linearly for 3D representations of avatars.** Both the pilot and main studies confirmed significant linear relationships between avatar density and emotional responses (valence and arousal) for 3D avatars, supporting the hypothesis.

Taken together, our findings indicate that increasing avatar density and different types of avatar representations lead to significant linear relationships with perceived crowding and emotional responses. The null hypothesis (**RQ4.H0**) is largely rejected, as significant linear relationships were found across multiple densities and measures (crowding perception, valence, and arousal).

Particle avatars generally result in less intense emotional responses compared to 2D and 3D avatars, supporting hypotheses **RQ4.H4** and **RQ4.H5**. The support for **RQ4.H6** is also strong, with 3D avatars showing linear variations in emotional responses with increasing density. Understanding these linear relationships is crucial for designing virtual environments that aim to maintain user comfort and positive experiences.

Chapter 6

Conclusions

In this thesis, I have explored the dynamics of crowding perceptions and emotional responses within virtual environments, focusing on the impact of avatar density and representation. This research extends the understanding of how digital spaces influence user experiences, particularly as our interactions increasingly transition to virtual platforms.

The investigation began with an exploration of consistency in user perceptions and emotional responses when exposed to varying crowd densities and different avatar representations (RQ1). The results revealed significant inconsistencies in users' crowding perceptions and emotional responses across multiple exposures, challenging the null hypothesis (RQ1.H0). The findings indicated that the type of avatar representation plays a critical role in influencing these perceptions, with 3D avatars generally leading to more consistent experiences compared to particle and 2D avatars. This insight underscores the need for careful consideration of avatar representation in the design of virtual environments to ensure stable user experiences.

The second research question (RQ2) sought to identify threshold densities at which users' emotional responses significantly change. The analysis revealed both common and unique thresholds across different avatar representations, partially rejecting the null hypothesis (RQ2.H0). The consistency of significant thresholds, particularly at density 8, across various measures highlights a fundamental aspect of how crowd density impacts user experience. These thresholds are crucial for designing virtual environments that mitigate negative emotional impacts and enhance user comfort.

Moving to the third research question (RQ3), the study examined whether different types of avatar representations and varying densities lead to differences in users' perceptions of crowding and emotional responses. The results consistently showed that different avatar representations significantly impact crowding perceptions and emotional responses, largely rejecting the null hypothesis (RQ3.H0). Particle avatars generally resulted in less intense emotional responses compared to 2D and 3D avatars, supporting hypotheses RQ3.H1 and RQ3.H2. However, the support for RQ3.H3 was partial, indicating that while 2D avatars sometimes led to less intense responses, 3D avatars did not consistently result in more intense emotional responses.

Finally, the fourth research question (RQ4) explored the linear relationship between avatar density and perceived crowding, and how emotional responses vary linearly with avatar density across different avatar representations. The findings revealed significant linear relationships between avatar density and both perceived crowding and emotional responses, largely rejecting the null hypothesis (RQ4.H0). The analysis confirmed that increasing avatar density leads to higher perceived crowding and varying emotional responses, with particle avatars generally resulting in less intense emotional responses compared to 2D and 3D avatars. These linear trends are essential for predicting user experiences and managing virtual spaces effectively.

In conclusion, this thesis provides insights into the emotional and psychological dynamics of virtual crowding. By studying various avatar types and densities, the research offers practical recommendations for designing more user-friendly virtual environments. The findings emphasize the importance of avatar representation in influencing user perceptions and emotional states, contributing to the broader understanding of crowding in both virtual and physical settings. This research bridges the gap between traditional architectural practice and virtual environments, ultimately aiming to enhance user comfort, satisfaction, and well-being in digital spaces. Future research can build on these insights to further refine virtual environment design and explore additional factors influencing user experiences in digital contexts.

6.1 Recommendations for future research

Although these studies support the understanding of how avatar density and representation affect user perceptions and emotional responses in virtual environments, their most important contribution may be that they raise a variety of questions for future study.

Future research could extend the current findings by examining the effect of repeated exposure over longer periods to determine if users' perceptions and emotions remain constant. Studying the influence of various user demographics (e.g., age, sex, cultural background) on consistency might offer deeper insights into personalized virtual environment design. Additionally, investigating how environmental elements like light and noise affect user response consistency is important.

Further work can refine this concept by examining additional physiological measures or emotional responses, providing a more complete picture of user experience. Understanding how users change their behavior in response to varying levels of density can be achieved by exploring changes in avatar numbers due to abrupt increases or decreases. Cross-validation for threshold densities across different virtual environments such as gaming, e-tourism, and online learning will improve generalizability.

It remains an open question until more work is done on the relationship between the extent of emotional response and other factors. Future research could seek nonlinear models to capture complex dynamics in user responses to varying avatar densities. Additionally, incorporating machine learning techniques would help increase prediction accuracy.

These directions aim to contribute to a deeper understanding of crowding perception within virtual reality settings and its effects on user experience.

6.2 Threats to validity and limitations

In this work, potential threats to validity could occur, which we have worked to address. For internal validity, repeated exposure to similar stimuli might result in participants learning how to respond or becoming fatigued, thus altering their responses and affecting the results. To mitigate this, the images were randomized to reduce the likelihood of learning effects, although we acknowledge that fatigue may still occur. Additionally, external events happening outside the experiment during the study period could influence participants' responses, introducing bias. We attempted to control for this by carefully monitoring the timing of the experiment. Changes in participants' behavior or responses due to the passage of time, unrelated to the experimental manipulations, also pose a threat. Furthermore, the combined influence of different variables, such as avatar density and representation, could complicate the interpretation of the results, making it difficult to isolate the effect of each variable. To address this, we employed a well-structured experimental design that carefully controlled and varied these factors systematically.

Regarding external validity, the sample used in the study may not be representative of the broader population, limiting the generalizability of the findings. We attempted to mitigate this by recruiting a diverse participant pool, although we recognize the inherent limitations of sample diversity. The study conditions may not accurately reflect real-world settings where virtual environments are typically experienced, which could limit the applicability of the findings to everyday situations. In terms of construct validity, other variables or constructs, such as personality traits and neurodiversity, may be inadvertently measured alongside the intended constructs, potentially confounding the results and making it unclear whether the outcomes are due to the intended manipulations or these extraneous factors. We addressed this by using validated measures to assess the primary constructs. Despite these efforts, some degree of construct confounding is unavoidable, and we have noted this limitation in our interpretation of the results.

In summary, while we have taken extensive measures to counteract potential threats to validity, some limitations remain inherent in our study design. By randomizing stimuli, monitoring external events, controlling for maturation, and carefully structuring our experimental design, we aimed to strengthen internal validity. To improve external validity, we recruited a diverse sample, though further replication is needed. For construct validity, we used validated measures, yet acknowledged the limitations of our analysis. These efforts reflect our commitment to conducting rigorous and reliable research while transparently addressing the challenges encountered.

Bibliography

[1] American Psychological Association. *Concise Guide to APA Style*. American Psychological Association, Washington, DC, 7th edition, 2020.

[2] Samaneh Aminikhanghahi and Diane Joyce Cook. A survey of methods for time series change point detection. *Knowledge and Information Systems*, 51:339 – 367, 2016.

[3] Dojin An and Jihang Lee. Effects of training using self-modeling with visual cues on skill performance, imagery, and sports confidence of adolescent female soccer players. *Korean Journal of Sport Science*, 2023.

[4] Chittaranjan Andrade. Z scores, standard scores, and composite test scores explained. *Indian Journal of Psychological Medicine*, 43:555 – 557, 2021.

[5] Arne Arnberger and Wolfgang Haider. A comparison of global and actual measures of perceived crowding of urban forest visitors. *Journal of Leisure Research*, 39:668 – 685, 2007.

[6] Arne Arnberger and Carsten Mann. Crowding in european forests: a review of recent research and implications for forest management and policy. *Forestry*, 81:559–571, 2008.

[7] Mousumi Banerjee, Michelle Capozzoli, Laura McSweeney, and Debajyoti Sinha. Beyond kappa: A review of interrater agreement measures. *Canadian Journal of Statistics*, 27, 1999.

[8] Madison Bentley, Christina Peerenboom, Frank Hodge, Edward B. Passano, Howard C. Warren, and Margaret Floy Washburn. Instructions in regard to preparation of manuscript. *Psychological Bulletin*, 26:57–63, 1929.

[9] Florian Berton, Ludovic Hoyet, Anne-Hélène Olivier, Julien Bruneau, Olivier Le Meur, and Julien Pettré. Eye-gaze activity in crowds: Impact of virtual reality and density. *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 322–331, 2020.

[10] Margaret M. Bradley, Maurizio Codispoti, Bruce N. Cuthbert, and Peter J. Lang. Emotion and motivation i: defensive and appetitive reactions in picture processing. *Emotion*, 1 3:276–98, 2001.

[11] Margaret M. Bradley and Peter J. Lang. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25 1:49–59, 1994.

[12] Marcelo Moll Brandão and Juracy Gomes Parente. Brasileiro gosta de "muvuca"?: impacto da densidade humana no comportamento de compra. *Rae-revista De Administracao De Empresas*, 52:613–627, 2012.

[13] Mohamad Adam Bujang and Nurakmal Baharum. A simplified guide to determination of sample size requirements for estimating the value of intraclass correlation coefficient: a review. 2017.

[14] Domenic V. Cicchetti and S A Sparrow. Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *American journal of mental deficiency*, 86 2:127–37, 1981.

[15] Luca Crociani, Lorenza Manenti, and Giuseppe Vizzari. Makksim: Mas-based crowd simulations for designer's decision support. In *Practical Applications of Agents and Multi-Agent Systems*, 2013.

[16] Jürgen Desor. Toward a psychological theory of crowding. *Journal of Personality and Social Psychology*, 21:79–83, 1972.

[17] Humeyra Dogru-Dastan. A chronological review on perceptions of crowding in tourism and recreation. *Tourism Recreation Research*, 47:190 – 210, 2020.

[18] Yiwen Dong, Peide Huang, and Hae Young Noh. Characterizing crowd preferences on stadium facilities through dynamic inverse reinforcement learning. *Proceedings of the 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, 2023.

[19] Anna Engelniederhammer, Georgios Papastefanou, and Luyao Xiang. Crowding density in urban environment and its effects on emotional responding of pedestrians: Using wearable device technology with sensors capturing proximity and psychophysiological emotion responses while walking in the street. *Journal of Human Behavior in the Social Environment*, 29:630 – 646, 2019.

[20] Fernando Enseñat-Soberanis and Rocío Blanco-Gregory. Crowding perception at the archaeological site of tulum, mexico: A key indicator for sustainable cultural tourism. *Land*, 2022.

[21] Gary W. Evans. The built environment and mental health. *Journal of Urban Health*, 80:536–555, 2003.

[22] Rupam Fedujwar and Amit Agarwal. A systematic review on crowding valuation in public transport. *Public Transport*, 2024.

[23] Victoria Filingeri, Ken Eason, Patrick Waterson, and Roger Haslam. Factors influencing experience in crowds – the participant perspective. *Applied Ergonomics*, 59:431–441, 2017.

[24] Paolo Fiorini and Zvi Shiller. Motion planning in dynamic environments using velocity obstacles. *The International Journal of Robotics Research*, 17:760 – 772, 1998.

[25] Joseph L. Fleiss. *Statistical methods for rates and proportions.* John Wiley & Sons, New York, 2nd edition, 1981.

[26] Hüseyin Gençer. Group dynamics and behaviour. *Universal Journal of Educational Research*, 2019.

[27] Joe A. Gliem and Rosemary R. Gliem. Calculating, interpreting, and reporting cronbach's alpha reliability coefficient for likert-type scales. 2003.

[28] Alan R. Graefe, Jerry J. Vaske, and Fred R. Kuss. Social carrying capacity: An integration and synthesis of twenty years of research. *Leisure Sciences*, 6:395–431, 1984.

[29] Stephen J. Guy, Jatin Chhugani, Sean Curtis, Pradeep K. Dubey, Ming C. Lin, and Dinesh Manocha. Pledestrians: a least-effort approach to crowd simulation. In *Symposium on Computer Animation*, 2010.

[30] Dirk Helbing, L'ubo Buzna, Anders Johansson, and Torsten Werner. Self-organized pedestrian crowd dynamics: Experiments, simulations, and design solutions. *Transp. Sci.*, 39:1–24, 2005.

[31] Dirk Helbing and Péter Molnár. Social force model for pedestrian dynamics. *Physical review. E, Statistical physics, plasmas, fluids, and related interdisciplinary topics*, 51 5:4282–4286, 1995.

[32] Seung Wan Hong, JiYoung Park, and Min-Jung Cho. Virtual vs. actual body: applicability of anthropomorphic avatars to enhance exploratory creativity in architectural design education. *Architectural Science Review*, 62:520 – 527, 2019.

[33] Ap Hudson-Smith, Martin Dodge, and Simon Doyle. Visual communication in urban planning and urban design. 1998.

[34] Motonari Isobe, Dirk Helbing, and Takashi Nagatani. Many-particle simulation of the evacuation process from a room without visibility. 2003.

[35] Rowena Jacobs, Maria Goddard, and Peter C. Smith. How robust are hospital ranks based on composite performance measures? *Medical Care*, 43:1177–1184, 2005.

[36] Philip Jelvard and Michael Mullins. The use of vr technologies to enhance methods for lighting design practice. 2019.

[37] Patty Kero and Daniel Lee. Likert is pronounced "lick-urt" not "lie-kurt" and the data are ordinal not interval. *Journal of applied measurement*, 17 4:502–509, 2016.

[38] Zinan Kiu. The dynamic effect of visual communication in web design and its technical realization. *Computer-Aided Design and Applications*, 2020.

[39] Sander Koelstra, Christian Mühl, M. Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and I. Patras. Deap: A database for emotion analysis ;using physiological signals. *IEEE Transactions on Affective Computing*, 3:18–31, 2012.

[40] Alexandros Koilias, Michael G. Nelson, Christos-Nikolaos E. Anagnostopoulos, and Christos Mousas. Immersive walking in a virtual crowd: The effects of the

density, speed, and direction of a virtual crowd on human movement behavior. *Computer Animation and Virtual Worlds*, 31, 2020.

[41] Terry K. Koo and Mae Y. Li. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2):155–163, 2016.

[42] Terry K Koo and Mae Y Li. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine*, 15 2:155–63, 2016.

[43] K. S. Anantha Krishna. Effect of architectural and spatial variables on the perception of crowding. *Architectural Science Review*, 34:95–108, 1991.

[44] Marina Lagune-Reutler, Andrew Guthrie, Yingling Fan, and David Matthew Levinson. Transit stop environments and waiting time perception: Impacts of trees, traffic exposure, and polluted air. *Transportation Research Record*, 2543:82 – 90, 2016.

[45] J Richard Landis and Gary G. Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33 1:159–74, 1977.

[46] Pj Lang. International affective picture system (iaps) : affective ratings of pictures and instruction manual. *CTIT technical reports series*, 2005.

[47] Zheng Li and David A. Hensher. Crowding in public transport: a review of objective and subjective measures. *The Journal of Public Transportation*, 16:6, 2013.

[48] Rensis Likert. A technique for the measurement of attitude scales. 1932.

[49] Alan Lipman and Edward Twitchell Hall. The hidden dimension. *British Journal of Sociology*, 21:353, 1970.

[50] Hongxiao Liu, Feng Li, Juanyong Li, and Yuyang Zhang. The relationships between urban parks, residents' physical activity, and mental health benefits: A case study from beijing, china. *Journal of environmental management*, 190:223–230, 2017.

[51] Wumei Liu, Zengguang Ma, and Xuhua Wei. A meta-analysis of the effect of crowding on consumers' emotional reactions and shopping-related behavioral reactions. *Acta Psychologica Sinica*, 2020.

[52] Jae Hoon Ma, Ece Erdogmus, and Seung Hyun Cha. Integration of a choice modeling approach with immersive virtual environments for accurate space utilization prediction. *Journal of Building Engineering*, 2023.

[53] Harry Mallgrave. *From Object to Experience: The New Culture of Architectural Design*. 01 2018.

[54] Shen Manqiong, Wang Haizhong, and Hu Guimei. A literature review and prospects of social crowding research in marketing. 2019.

[55] Uwe Messer and Alexander Leischnig. Social crowding and individuals' risk taking: A life-history theory perspective. 2018.

[56] Ali Mollahosseini, Behzad Hasani, and Mohammad H. Mahoor. Affectnet: A database for facial expression, valence, and arousal computing in the wild. *IEEE Transactions on Affective Computing*, 10:18–31, 2017.

[57] Matthias Müller and Nuttapong Chentanez. Solid simulation with oriented particles. *ACM SIGGRAPH 2011 papers*, 2011.

[58] Soraia Raupp Musse and Daniel Thalmann. A model of human crowd behavior : Group inter-relationship and collision detection analysis. In *Computer Animation and Simulation*, 1997.

[59] National Quality Forum. Composite performance measure evaluation guidance. https://www.qualityforum.org/Projects/n-r/Performance$_{M}easure_{E}valuation_{F}ramework/Composite_{E}valuation_{G}uidance.aspx, April201$ $2024-07-12$.

[60] Bart Neuts and Dominique Vanneste. Contextual effects on crowding perception: An analysis of antwerp and amsterdam. *Tijdschrift voor economische en sociale geografie*, 109:402–419, 2018.

[61] Geoff Norman. Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education*, 15:625–632, 2010.

[62] Bushra Obeidat and Esra'a Abdul Rahman Jaradat. The influence of virtual human representations on first-year architecture students' perceptions of digitally designed spaces: a pilot study. *Building Research & Information*, 50:845 – 858, 2022.

[63] Agnieszka Olszanecka, Wiktoria Wojciechowska, Agnieszka Bednarek, Piotr Kusak, Barbara Wizner, Michał Terlecki, Katarzyna Stolarz-Skrzypek, Marek Klocek, Tomasz Drożdż, Krzysztof Sładek, Monika Bociaga-Jasik, Aleksander Garlicki, Krzysztof Rewiuk, Andrzej Matyja, Maciej Małecki, Wojciech Sydor, Marcin Krzanowski, Tomasz Grodzicki, and Marek W. Rajzer. Serial echocardiographic evaluation of covid-19 patients without prior history of structural heart disease: a 1-year follow-up cracov-hhs study. *Frontiers in Cardiovascular Medicine*, 10, 2023.

[64] Christopher J. Patrick and Stacey A. Lavoro. Ratings of emotional response to pictorial stimuli: Positive and negative affect dimensions. *Motivation and Emotion*, 21:297–321, 1997.

[65] Ferdinand Pittino, Lisa Valentina Eberhardt, Anja Kurz, and Anke Huckauf. Crowding with negatively conditioned flankers and targets. *Advances in Cognitive Psychology*, 15:1 – 10, 2019.

[66] Judith Rodin. Density, perceived choice, and response to controllable and uncontrollable outcomes. *Journal of Experimental Social Psychology*, 12:564–578, 1976.

[67] Silvia Sanz-Blas, Daniela Buzova, and Fernando J. Garrigós-Simón. Understanding crowding perceptions and their impact on place experience: Insights from a mixed-methods study. *Psychology & Marketing*, 2024.

[68] Patrick Schober, Christa Boer, and Lothar A. Schwarte. Correlation coefficients: Appropriate use and interpretation. *Anesthesia & Analgesia*, 126:1763–1768, 2018.

[69] Mathew Schwartz. From human inspired design to human based design. 2017.

[70] Mathew Schwartz. Human centric accessibility graph for environment analysis. *ArXiv*, abs/2101.10503, 2021.

[71] Mathew Schwartz, Brandon Haworth, Muhammad Usman, Petros Faloutsos, and Mubbasir Kapadia. Impact of manikin display on perception of spatial planning. *ACM Symposium on Applied Perception 2022*, 2022.

[72] Zahra Shahhoseini and Majid Sarvi. Traffic flow of merging pedestrian crowds: How architectural design affects collective movement efficiency. *Transportation Research Record*, 2672:121 – 132, 2018.

[73] Michael R. Sheldon, Michael J. Fillyaw, and W. Douglas Thompson. The use and interpretation of the friedman test in the analysis of ordinal-scale data in repeated measures designs. *Physiotherapy research international : the journal for researchers and clinicians in physical therapy*, 1 4:221–8, 1996.

[74] Patrick E. Shrout and Joseph L. Fleiss. Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin*, 86 2:420–8, 1979.

[75] Michael Shwartz, Joseph D. Restuccia, and Amy K Rosen. Composite measures of health care provider performance: A description of approaches. *The Milbank quarterly*, 93 4:788–825, 2015.

[76] Robert Sommer. Personal space : the behavioral basis of design. *American Sociological Review*, 35:164, 1970.

[77] Daniel Stokols. On the distinction between density and crowding: some implications for future research. *Psychological review*, 79 3:275–7, 1972.

[78] Daniel Stokols. A social-psychological model of human crowding phenomena. *Journal of The American Planning Association*, 38:72–83, 1972.

[79] Ya-Yen Sun and Megha Budruk. The moderating effect of nationality on crowding perception, its antecedents, and coping behaviours: A study of an urban heritage site in taiwan. *Current Issues in Tourism*, 20:1246 – 1264, 2017.

[80] Debby ten Hove, Terrence D. Jorgensen, and L. Andries van der Ark. Updated guidelines on selecting an intraclass correlation coefficient for interrater reliability, with applications to incomplete observational designs. *Psychological methods*, 2022.

[81] Daniel Thalmann. Crowd simulation. In *Wiley Encyclopedia of Computer Science and Engineering*, 2007.

[82] Mingliang Xu, Hao Jiang, Xiaogang Jin, and Zhigang Deng. Crowd simulation and its applications: Recent advances. *Journal of Computer Science and Technology*, 29:799 – 811, 2014.

[83] Cindy Xue, Jing Yuan, Gladys Goh Lo, Amy Tien Yee Chang, D.M.C. Poon, Oi Lei Wong, Yihang Zhou, and Winnie Chiu Wing Chu. Radiomics feature reliability assessed by intraclass correlation coefficient: a systematic review. *Quantitative imaging in medicine and surgery*, 11 10:4431–4460, 2021.

[84] Catherine A. Zanbaka, Amy Banic, Paula Goolkasian, and Larry F. Hodges. Social responses to virtual humans: implications for future interface design. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2007.

[85] Anita Zehrer and Frieda Raich. The impact of perceived crowding on customer satisfaction. *Journal of Hospitality and Tourism Management*, 29:88–98, 2016.

[86] Julius Panero Martin Zelnik. Human dimension & interior space: A source book of design reference standards. 1979.

[87] Jingjing Zhang, Thammathip Piumsomboon, Ze Dong, Xiaoliang Bai, Simon Hoermann, and Robert W. Lindeman. Exploring spatial scale perception in immersive virtual reality for risk assessment in interior design. *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020.

[88] Katja Zibrek, Seán J. Martin, and Rachel Mcdonnell. Is photorealism important for perception of expressive virtual humans in virtual reality? *ACM Transactions on Applied Perception (TAP)*, 16:1 – 19, 2019.

[89] Donald W. Zimmerman and Bruno D. Zumbo. Relative power of the wilcoxon test, the friedman test, and repeated-measures anova on ranks. *Journal of Experimental Education*, 62:75–86, 1993.