At the Metro Entrance Realm

How lighting and trees shape stress at Bullewijk, a Virtual Reality study

Master of Science Thesis

Written by Yuval Tshuva

Metropolitan Analysis, Design, and Engineering
Delft University of Technology, Wageningen University & Research

Supervised by Prof. Alexander Klippel, Dr. Yan Feng, Dr. Jaime Soza Parra

November 2025

TUDelft AMS AMSTERDAM INSTITUTE FOR ADVANCED METROPOLITAN SOLUTIONS MSc MADE WAGENINGEN UNIVERSITY & RESEARCH

Author

Yuval Tshuva

WUR student number: 1271105, TU Delft student number: 5860342

Academic Supervisors

Prof. Dr. Alexander Klippel
Wageningen University of Technology
Cultural Geography

Dr. Yan Feng
Delft University of Technology
Faculty of Civil Engineering, Transport & Planning

Content Supervisor

Dr. Ing. Jaime Soza Parra
Utrecht University
Geosciences, Human Geography and Spatial Planning

Amsterdam Institute for Advanced Metropolitan Solutions (AMS)
Amsterdam, The Netherlands
November 2025

Abstract

This thesis examines how lighting and trees influence experienced stress in a metro station plaza. A Virtual Reality experiment of a site-specific environment at the Bullewijk station plaza in Amsterdam was conducted with 47 participants. Each participant experienced five scenarios that varied in day or night, lighting tone and distribution, and the presence of trees, including the current as-is configuration. Stress was assessed using questionnaires (SSSQ-10 and single-item appraisals of safety, attractiveness, and intended use) and an Empatica E4 wristband to measure electrodermal activity and heart rate. Statistical analysis using Linear mixed models, cumulative link mixed models, and planned contrasts, showed clearer psychological than physiological effects. Night scenes increased distress and reduced perceived safety and attractiveness relative to day scenes. At night, warm lighting increased safety, attractiveness, and the intended use compared with the existing cool lighting. Trees added smaller but positive contributions. Women showed larger night penalties and stronger gains from warm lighting, while age-related differences were not found to be a strong predictor. Interpreted through ethical-spatial and behavioral lenses, the findings suggest that upgrades in night lighting, and possible addition of trees, can reduce experienced stress at station plazas, enhance the quality of public spaces, support inclusivity, and improve pedestrian mobility opportunities.

Keywords: Virtual Reality, Urban design, Nighttime lighting and trees, Perceived stress, Metro station plaza.

Table of Contents

List of abbreviations

BH FDR	Benjamini-Hochberg false discovery rate
BVP	Blood volume pulse
CCT	Correlated color temperature
CLMM	Cumulative link mixed model
EDA	Electrodermal activity
HR	Heart rate
HRV	Heart rate variability
iPQ	igroup Presence Questionnaire
LMM	Linear mixed model
OR	Odds ratio
PO	Proportional odds
PSS 4	Perceived Stress Scale, 4-item version
QC	Quality control
SCL	Skin conductance level
SCR	Skin conductance response
SDNN	Standard deviation of normal to normal intervals
SRQ	Sub research question
SSQ	Simulator Sickness Questionnaire
SSSQ	Short State Stress Questionnaire
VR	Virtual reality

Glossary

Stress	Short-term response to environmental demands that are evaluated as taxing or exceeding one's resources and endangering well-being. This thesis focuses on acute stress during brief VR scene exposures.
Perceived stress	The subjective experience of stress during or after a scene, inferred from self-reported questionnaires (Distress and Engagement scales and single-item ratings of Safety, Attractiveness, and Intended Use).
Channels	The mixed-methods nature of this research includes psychological and physiological outcomes, which are collectively referred to as channels in this paper.

Physiological measurements	Sensor-based indicators of autonomic activation, recorded with the Empatica E4 wristband, specifically Skin Conductance Level (SCL) and Heart Rate (HR).
Psychological questionnaires	Self-report instruments that capture momentary stress states and appraisals of the place, mainly the Short State Stress Questionnaire (SSSQ) and single-item ratings of Safety, Attractiveness, and Intended Use.
Distress	SSSQ subscale reflecting unpleasant arousal, tension, and perceived overload. Used as the primary subjective indicator of stress.
Engagement	SSSQ subscale reflecting focused attention, interest, and energetic involvement. Captures motivated arousal that is not necessarily negative.
Psychological appraisals	Scene-specific self-reported ratings of safety, attractiveness, and intended use of the plaza.
Physiological arousal indices	Skin Conductance Level (SCL) and Heart Rate (HR) per scenario, used as body-based indicators of stress-related autonomic activation.
Station plaza	The public space directly outside the metro entrance is the entrance realm. In this thesis, the plaza at Bullewijk station forms the threshold between street and transit.

List of figures

List of tables

All photographs, figures, and visualizations in this thesis were created by the author, unless stated otherwise.

1 Introduction

1.1 Background

Public space and mobility patterns are key determinants of urban quality and vitality (Schönfeld & Bertolini, 2017). Global urbanization trends have led to cities to rapidly expand with new residential districts that often prioritize rapid delivery and low cost, featuring wide streets, generous parking, and greater building setbacks (Ritchie et al., 2024). These districts may include under-designed public spaces, with extensive paving, limited or monotonous greenery, harsh lighting, and little to invite lingering (Knaap & Talen, 2005; Savini et al., 2016; Schönfeld & Bertolini, 2017). Station plazas, the spaces immediately outside metro entrances, are a salient example of underwhelming design (Savini et al., 2016). Modern urban expansion has frequently produced entrance realms that serve movement efficiently but offer little in terms of human experience (Hamilton-Baillie, 2008). In Amsterdam, many metro entrances sit under rail bridges, creating long underpasses that can feel exposed or unsafe (Schönfeld & Bertolini, 2017). These plazas serve as the threshold to public transportation and are encountered by most pedestrians in brief, transitional moments that are likely to shape who chooses to use the system (Rossetti et al., 2019). Some encounter these places unpleasantly, which leads to negative perceptions of safety and attractiveness and can be expressed as experienced stress. Small design decisions, such as lighting and the addition of trees, can shape the experience in public spaces, particularly at night (Chen et al., 2024).

As European cities seek to reduce car travel and improve public transportation, this ambition depends on public mobility that feels comfortable and inclusive (Savini et al., 2016). Station plazas are critical because they determine whether a passenger decides to cross over to the station or avoid it (Du et al., 2021; Walker, 2024). However, these plazas are situated between studied categories, such as streets, parks, and metro interiors, and receive less attention in empirical research on stress and perceptions of public space. Specific findings are essential for transit agencies and municipalities to improve these places.

1.2 Context

At the threshold of transit, the perceived stressfulness of a public space affects both inclusivity and mobility. If the threshold to transit feels exposed, tense, or unsafe, some people will avoid using it, especially at night, which undermines the value of equal access to the city. Bullewijk metro station plaza is a good place to study this. It is a relatively simple and underused entrance realm, with little programming or shelter and long views under the bridge, so small changes can shift how it feels.

Virtual Reality (VR) enables testing these changes without altering the actual site. In this thesis, the Bullewijk station plaza is reconstructed as a site-specific virtual environment in which lighting and tree configurations can be varied while the rest of the setting remains the same. This allows the plaza to be studied as a place where stress is shaped by design, without actually interfering. A detailed description of Bullewijk and the station is provided in Chapter 3.

1.3 Current state of the literature

This section reviews four domains of literature that inform the sub-research questions (SRQ1-SRQ4), hypotheses (H1-H4), and the design of the VR scenarios. The first domain treats stress in public space and motivates a focus on station plazas as threshold spaces for transit. The second and third domains examine how nighttime lighting and trees influence stress. The fourth addresses how gender and age shape stress at night, motivating the planned moderation analysis. Together, these domains provide the empirical basis for SRQ1-4 and H1-4 and are schematically mapped in Figure 1.1.

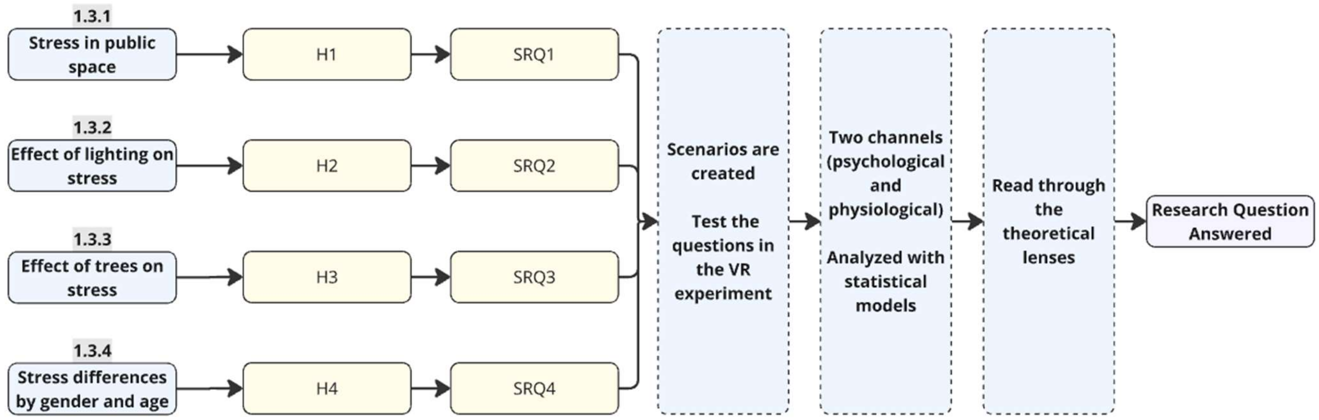

Figure 1.1: Mapping of the literature review to the hypotheses, sub-research questions, and the pipeline towards answering the research question.

1.3.1 Stress in public space

Rapid urbanization has introduced multiple environmental and social stressors: high population density, pollution, noise, economic pressure. These factors contribute to psychological and physiological stress and pose a substantial threat to the well-being of urban residents (Li & Lange, 2023). Urban expansion has been associated with elevated cortisol levels, anxiety, and other stress-related health risks, making stress in cities a growing public health concern (Hedblom et al., 2019; Li & Lange, 2023).

In transit environments, perceived safety is a primary determinant of whether pedestrians are willing to use a space. Unfamiliar or unconventional street layouts tend to be perceived as less safe and elicit measurably higher stress responses (Argota Sanchez-Vaquerizo et al., 2023). Station plazas are a specific type of public space where this stress-safety link is critical: they are brief but critical thresholds where travelers decide whether to approach and use public transport, especially at night. Inadequate design at this threshold can therefore undermine equitable access to transit (White et al., 2023; Yang et al., 2024).

Beyond single elements, the overall quality of public space helps determine whether it feels pleasant. Gehl's human scale quality criteria emphasize that inclusive public spaces should be well lit at pedestrian scale, protect users from unpleasant glare and provide unhindered views for orientation (Gehl, 2013). Taken together, this literature suggests that in threshold spaces such as station plazas, stress is shaped by a combination of urban stressors, perceived safety and human scale design features. Lighting, visibility, and the presence of greenery interact to influence whether the plaza feels like a controllable, usable space or a tense environment to be avoided. This underpins the focus of this thesis on the Bullewijk station plaza as a transit threshold and motivates the experimental manipulation of nighttime lighting and trees in a controlled VR setting (Brambilla et al., 2024; White et al., 2023). See section 2.3 for the definition of stress in this experiment.

1.3.2 Effect of lighting on stress

Lighting is a critical environmental feature that modulates stress and perceived safety in public spaces, particularly at night. Higher illuminance levels and greater lighting uniformity are consistently associated with lower stress responses and higher perceived safety (Rabaza et al., 2013). Effective nighttime design must balance visibility, energy efficiency, and reassurance. This is crucial during winter, when nights are longer, and some commute twice a day in the dark (Scorpio et al., 2020).

The quality of light is defined by parameters such as Correlated Color Temperature (CCT) and Spectral Power Distribution (SPD), and influences subjective appraisal. Scorpio et al. claims that warm light (lower CCT) is generally linked to comfort and positive mood, while cool light (higher CCT) may enhance facial recognition and, consequently, perceived safety (Scorpio et al., 2020). The qua

Importantly, the distribution of light must prevent negative visual experiences, such as glare and deep shadows, which significantly undermine visibility and heighten fear perception (Park et al., 2025; Scorpio et al., 2020). VR and recorded environment experiments show that when nighttime illuminance is low, users report higher fear and negative emotions, whereas increasing illuminance and ensuring appropriate distributions of artificial light can significantly reduce fear of crime in narrow streets and residential areas (Park et al., 2025; Son et al., 2024).

1.3.3 Effect of trees on stress

The quality of public space strongly influences the psychological and physiological stress experienced by urban residents. High levels of urbanization introduce stressors such as noise, pollution, and crowding, which are linked to adverse health outcomes, whereas access to urban green spaces can help buffer these effects (Kardan et al., 2016; Tyrväinen et al., 2014). Empirical studies consistently find that closeness to nature is associated with decreased stress, anxiety, anger, and fatigue, as well as better cognitive performance and increased happiness (Llaguno-Munitxa et al., 2022; Twohig-Bennett & Jones, 2018). The combined effects of night lighting and greenery have a strong influence on the utility and perceived quality of urban spaces. Greenery, even in virtual environments, has a measurable restorative effect, reducing stress and increasing positive mood (Bertram & Rehdanz, 2015)

Trees introduce design trade-offs. Vegetation structure, openness, and lighting strongly influence whether a space feels like a safe refuge or a place of potential danger (Rossetti et al., 2019; Wan et al., 2024; Yu et al., 2018; Zhang et al., 2022). Poorly placed trees or poles can create dense shadows and hidden corners, undermining clear sightlines and compromising perceived safety, especially at night (Scorpio et al., 2020). In station plazas, this means trees can soften the visual environment and enhance restoration, but only when integrated with appropriate lighting and visibility. These insights motivate the experimental manipulation of trees in combination with nighttime lighting in the VR station scenarios, to test when greenery supports stress recovery and when it may inadvertently heighten fear. A Recent VR and immersive study extend these findings to digitally rendered environments (Hedblom et al., 2019). Adding green elements, such as plants or green walls, reduces perceived stress, improves self-reported well-being, and can enhance some aspects of cognitive performance.

1.3.4 Stress differences by gender and age

Perceived stress and feelings of unsafety in public spaces are not evenly distributed across social groups. At night, women consistently report higher levels of fear and perceived unsafety than men across a range of urban contexts, with these differences often linked to lighting quality, visibility, and social cues in the environment (Fotios et al., 2015; Johansson & Haandrikman, 2023). Studies on pedestrian reassurance show that appropriate road lighting, providing sufficient illuminance, uniformity, and facial recognition, can reduce fear after dark, with stronger benefits frequently observed for women (Fotios et al., 2015). Built environment features such as openness, clear sightlines, and the presence of other people similarly influence perceived safety, particularly for women and other groups at higher risk of harassment. Du et al. find that women place more importance on safety and basic needs and wayfinding at Amsterdam Centraal. Their survey-based work points on gender sensitivity to safety (Du et al., 2021).

Age is another important factor. Older adults are generally more sensitive to low illuminance and glare due to age-related vision changes, which can increase stress, disorientation, and avoidance of public spaces at night (Zhang et al., 2022). Maintaining uniform illumination, minimizing glare, and ensuring clear sightlines are therefore especially critical when designing nighttime environments for an aging population. Recent simulation and VR studies reinforce these patterns: as ambient light decreases, fear and stress responses tend to rise, and these effects are moderated by gender and age, with women and older participants often reporting greater discomfort under darker or less uniform lighting conditions (Argota Sanchez-Vaquerizo et al., 2023; Scorpio et al., 2020). These findings claim that nighttime stress in public space is socially differentiated and that design interventions, particularly in lighting and visibility, are unlikely to be experienced uniformly. They motivate the planned moderation by gender and age in the analysis of the VR experiment and justify a specific focus on lighting tone, distribution, and interactions with trees to target reductions in stress in station plazas.

1.4 Research Gaps and Problem Statement

While public space perception is commonly studied, several gaps in the existing work motivated this study. First, station plazas at metro entrances appear to sit between well-defined categories such as streets, parks, and station interiors, and therefore receive comparatively less direct attention in design and research. Much of the literature focuses either on the station building and platforms or on streets and parks, rather than on the plaza outside the entrance, where many access decisions are actually made (Rossetti et al., 2019; Savini et al., 2016; Schönfeld & Bertolini, 2017).

Second, VR has been used to study pedestrian behavior, safety, and stress. However, many VR studies are done in generic or stylized environments rather than in site-specific virtual environments that replicate an existing location, which makes it difficult to transform findings into design guidelines in accordance with unique geometry, lighting constraints, and social context (Feng et al., 2021; Ghanbari et al., 2024; Scorpio et al., 2020; White et al., 2023)

Third, there is limited empirical research on the experience of Amsterdam's metro station areas in Amsterdam Zuidoost, despite their importance for access to jobs and services. Evidence from such locations could support more locally grounded decisions about everyday public space and transit access.

There are some works on Amsterdam's mobility and perception, but less on the metro, and especially not on Zuidoost (Du et al., 2021; Savini et al., 2016; Schönfeld & Bertolini, 2017).

Fourth, many VR studies that include physiological measures restrict participants to standing still or moving via teleportation. This improves signal quality but departs from everyday walking behavior in public space. Conversely, allowing natural walking in short exposures, as in this study, increases ecological realism but makes some arousal metrics hard to use reliably. There is therefore a need for designs that balance natural movement with robust stress-related indices (Bzdúšková et al., 2022; Feng et al., 2021; Ghanbari et al., 2024; Hedblom et al., 2019; Park et al., 2025; Son et al., 2024).

This thesis addresses these gaps by holding the place constant and manipulating lighting and trees in an immersive virtual environment of the Bullewijk metro station plaza, and then measuring changes in participants' stress levels. The core problem is whether feasible adjustments to a real station plaza can measurably change how the space is experienced and how willing people are to use the metro it serves.

This thesis is written as a part of the Metropolitan Analysis, Design and Engineering (MADE) program at the AMS Institute. It contributes to the institute's metropolitan challenge of Smart Urban Mobility by examining how station plazas can support a low-car city vision and reduce dependence on private automobiles by making access to the metro more inclusive, safe, and attractive, and by strengthening the station area as a functional transportation hub that begins in the plaza itself.

1.5 Research Objectives

This study focuses on people who can travel by metro and experience the public space surrounding as pedestrians. Its goal is to support sustainable, inclusive, low-car mobility by improving the quality of the pedestrian environment at station plazas. Specifically, this VR site-specific experiment aims to understand how lighting and trees at the plaza influence travelers through two stress-related channels, psychological self-reports and physiological responses. The study uses within-participant comparisons while holding other attributes of the plaza constant under controlled, repeatable lab conditions. Lastly, it seeks to link public space design and transit access, and to provide design guidance for improving similar station plazas. This can support inclusive mobility and offer better opportunities for pedestrians.

1.6 Research Questions and Hypotheses

The main research question is framed to capture the overall experience of the public space at the station threshold within the theoretical framework and the VR evaluation. The main research question is:

How can adjustments in station-plaza design, specifically nighttime lighting and the presence of trees, affect pedestrians' experienced stress at transit stations?

The sub-research questions further elaborate on the factors that contribute to the result, specifically those relevant to the Bullewijk metro station plaza. The Sub-Research Questions (SRQ) are:

SRQ1: How does pedestrians' stress differ between day and night variants of the plaza?
SRQ2: At night, how do changes in lighting attributes change stress, with and without trees present?

SRQ3: How does the presence of trees change stress across day and night?
SRQ4: Does the impact of lighting and trees on stress differ by gender and age?

The following hypotheses help answer the sub-research questions. They align with the literature domains reviewed in Sections 1.3.1-1.3.4 and with SRQ1-SRQ4. They express expected directions of effect and trends on stress levels, and guide the interpretations of the results in the discussion chapter.

H1: Pedestrians will experience higher stress in nighttime scenarios compared to daytime.
H2: Warmer and uniform lighting reduced stress compared with cooler baseline as-is patched lighting.
H3: More trees in the plaza reduce stress compared to baseline as-is with effects that may depend on time of day.
H4: Gender and age differences will incur in significant experienced stress throughout the scenarios.

1.7 Scope of the Thesis

The study focuses on the specific site in Bullewijk. A virtual environment of this plaza was created, and participants experienced five scenarios within a single session. The exposure window for each scenario lasted 90 seconds. Forty-seven participants were recruited, and data were collected using physiological measures (an Empatica E4 wristband, skin conductance, and heart rate) and psychological measures (questionnaires). The study examines the moderating effects of gender and age. It does not measure ridership and long-term behavior changes, does not focus on crime outcomes, and does not generalize beyond sites with similar spatial and lighting conditions.

Thesis structure. The structure of this thesis is as follows: Chapter 2 provides an overview of the framework of this paper, reviewing the theoretical lens guiding the analysis and operationalization of experienced stress in the context of the experiment. Chapter 3 describes the urban context of the Bullewijk metro station and the VR environment, and follows the adjustments identified in the site analysis, the experimental design, and the measures. Chapter 4 elaborates on the quality control rules, statistical analysis guides, models, and formulations, and ethical concerns of the experiment. Chapter 5 presents the psychological and physiological outcomes of the experiment. Chapter 6 discusses the implications for design and policy, the key findings in accordance with the theoretical lenses, limitations, and directions for further research and recommendations. Chapter 7 concludes this thesis.

2 Theoretical Framework

This chapter introduces the theoretical lenses used to interpret the experiment and specifies how stress and related perceptions are understood in this thesis. To explain why these changes matter and how to read the results, the study uses three complementary lenses. The capabilities lens provides the overarching ethical perspective: does the plaza expand people's real freedom to use transit without excessive stress? Civility and design qualities translate that aim into eye-level qualities of a respectful, legible, comfortable entrance. The theory of planned behavior is used to understand the scope of this experiment in terms of long-run behavior and immediate intention. Together, these lenses guide interpretation, justify the need for such a study at this location, define the design interventions, and suggest critical lenses of reading the results. The final part of this chapter presents how stress is conceptualized through psychological questionnaires and physiological measurements.

2.1 Ethical-Spatial Lenses: Capabilities Approach

The Capabilities Approach is a framework that evaluates individual wellbeing and social justice by focusing on what people are actually able to be and do. That is contrary to focusing on resources or subjective utility. Sen's theory (Sen, 1999, 2008) suggests that situations are judged by the *real freedoms* people have to do and be what they value. The approach distinguishes functionings, the actual beings and doings people realize, such as using the metro, from capabilities, the opportunities to achieve those functionings. The same nominal resources can yield different real freedoms because of conversion factors that mediate how resources translate into capabilities. These include personal factors such as gender and age, social factors such as the local safety climate, and environmental factors such as design, lighting, and trees.

Nussbaum's list of central capabilities helps identify which domains are at stake in specific contexts (M. Nussbaum, 2003; M. C. Nussbaum, 2005). At the station entrance and plaza, the most relevant are bodily integrity, the ability to move without fear, and control over one's environment and the ability to act in public space. Robeyns provides operational guidance (Robeyns, 2005; Robeyns & Byskov, 2011) that makes the approach usable in empirical studies. This involves specifying a focal capability, identifying relevant conversion factors, the personality in a social context, and selecting transparent indicators that can track changes in capabilities in a given context. Robeyns also stresses that mobility is itself a valuable freedom, not only a means to other ends. This critiques traditional approaches that focus on resources or on how far people can travel to measure accessibility. Building on this, Vecchio and Martens' concept of Accessibility as a Capability (Vecchio & Martens, 2021) reframes accessibility around what people can actually do, the activity opportunities they can realistically reach, and how private and public mobility resources interact with these opportunities. This perspective explains why expanding the capability to the plaza matters for spatial justice even when network resources are fixed.

This thesis applies the Capabilities Approach to evaluate the Bullewijk station plaza as a public space. The focal capability is defined as being able to access and use the station plaza and metro, especially at night, without fear or undue stress. Environmental conversion factors are the design features tested in this study, namely, lighting and trees in the station plaza, addressed in SRQ1-3. Personal conversion factors, gender, and age are modeled as moderators in SRQ4. The wider social context, such as the neighborhood's safety climate, is acknowledged but not manipulated. These elements correspond to the allocation of resources, conversion factors, capabilities, functionings, and agency in Figure 2.1.

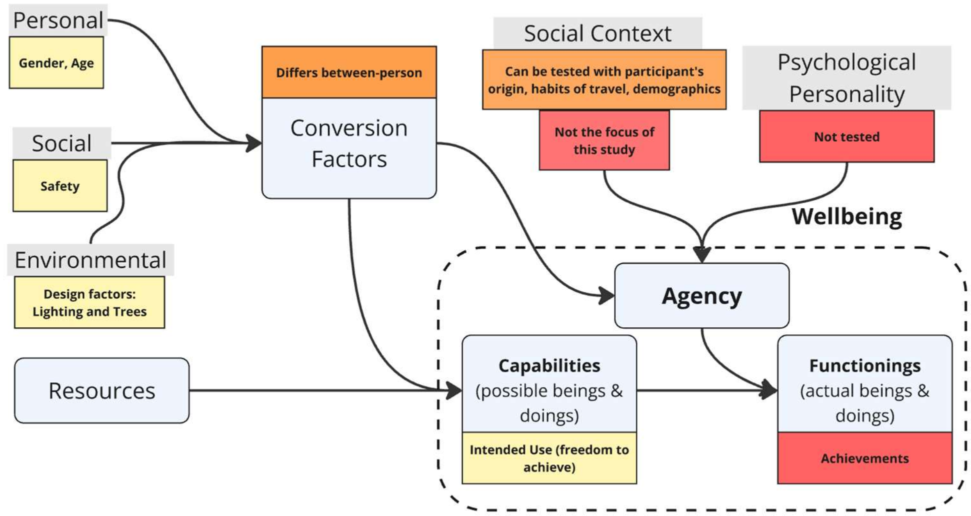

Figure 2.1: Capabilities Approach applied to this thesis. In yellow are the constructs for this thesis, red is the limitation, and orange are limitations that are partially addressed.

2.2 Behavioral Lenses: Theory of Planned Behavior, Criteria of Public Space

Ajzen's theory of planned behavior (TPB) (Ajzen, 1991) links beliefs to behavior through intention. Intention to perform a behavior is shaped by three components: attitude toward performing the behavior, subjective norm of how others behave, and perceived behavioral control (PBC), the individual's perception of the ability to control the behavior. These component rests on individual's beliefs. They lead to the intention, the willingness to perform an action, which dictates the behavior, the observable response. In this thesis, TPB is used to connect design and perception: attitude is represented by perceived safety and attractiveness, PBC is partly shaped by design features in the plaza, and intention is captured as intended use in VR. Behavior change, such as long-run ridership or actual station use, is not observed (Figure 2.2).

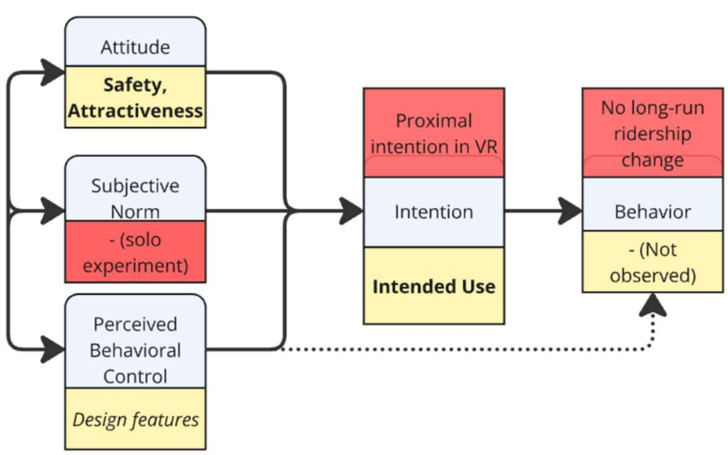

Figure 2.2: Theory of Planned Behavior, applied to this thesis. In yellow are the constructs, and red are the limitations.

Gehl's human-scale criteria specify 12 qualities of public space (Gehl, 2013) grouped into three requirement domains. Protection concerns feeling safe, secure, and sheltered, for example being shielded from traffic conflicts and having lighting cover from sun, rain, noise, and glare. Comfort covers basic activities, including opportunities to walk, stay, sit, see, talk and listen, and play and exercise. Enjoyment relates to human-scale dimensions and aesthetic qualities that make a place pleasant to spend time in. Walker's seven demands of a useful transit system extend this to transit-oriented public spaces (Walker, 2012, 2024) Two demands are especially relevant at the station threshold: "It respects me," referring to the physical environment and the social experience of using it, and "It is a good use of my time," referring to the quality of the station as a node in a trip chain. Together, these are summarized as civility, including safety, security, amenity, courtesy, and cleanliness (Figure 2.3).

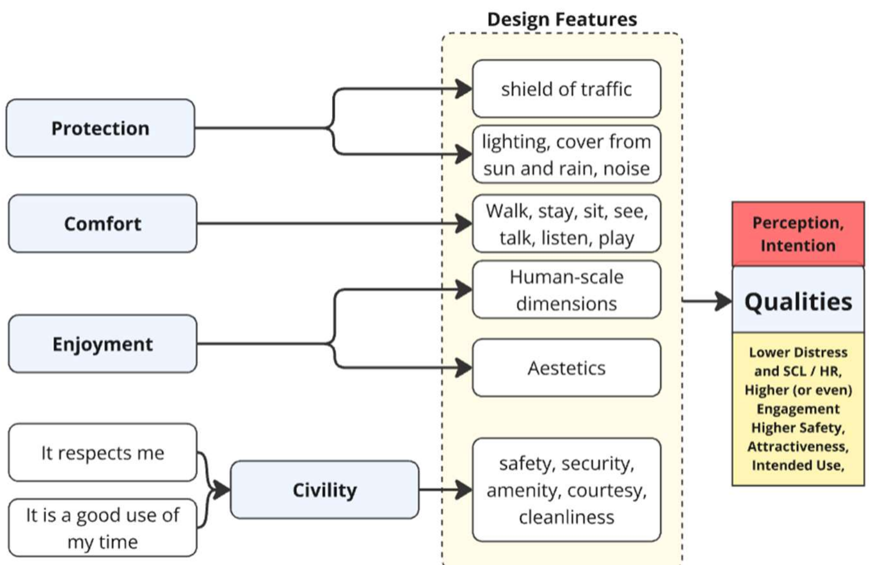

Figure 2.3: Behavioral lenses of twelve qualities (Gehl, 2013) and seven demands (Walker, 2011). In yellow are the constructs, and red are the limitations.

2.3 Stress

Stress can be defined as a state of worry or mental tension. It is a natural human response that prompts people to address challenges and threats in their lives (World Health Organization, 2023). Acute stress is a short aversive state triggered by a specific stimulus that disrupts homeostasis, which means it produces mental tension, physiological arousal, and a feeling of overload (Epel et al., 2018). A Notable stress theory is the Transactional Model of Stress and Coping (Biggs et al., 2017; Lazarus & Folkman, 1984), which claims that stress results from an appraisal process, followed by bodily reactions that can be measured both subjectively and objectively. Self-reported scales provide insights into perceived stress and threat appraisal (Bzdúšková et al., 2022; Yang et al., 2024), while physiological responses are tested to measure stress in VR research (Lazarus & Folkman, 1984; Ma et al., 2025). Together, these measures allow a comprehensive assessment of how lighting and tree configurations around Bullewijk metro station influence pedestrians' acute stress responses and their sense of safety.

3 Study Site and Virtual Reality

This chapter outlines the design of the virtual reality (VR) experiment and the study site. It first describes the research design and mixed methods approach, then introduces the Bullewijk metro station plaza and its main spatial characteristics. This is followed by a description of the construction of the virtual environment in the Unreal Engine ecosystem, the experimental apparatus and room settings, and the recruitment and participant characteristics. Chapter 4 then reports the analysis plan for the physiological and psychological data and elaborates on the statistical models applied.

3.1 Research Design

The study uses a site-specific VR experiment with a mixed-methods design that combines psychological self-report questionnaires and physiological stress measurement responses. The experiment uses a within-participant design, with each participant experiencing five variations of the same station plaza in Amsterdam. The scenarios differ in time of day (day versus night), lighting tone and distribution, and the presence or absence of trees, while all other features of the plaza remain constant (see Section 3.3).

In each session, participants walk through all five scenarios in randomized order using an immersive head-mounted display. Stress is measured per scenario through two channels. The psychological channel is based on the questionnaires and provides appraisals, distress, and engagement scales. The physiological channel consists of arousal indices derived from an Empatica E4 wristband, primarily skin conductance level (SCL) and heart rate (HR) recorded continuously during each exposure. Questionnaires are administered immediately after every scene, aligned with the physiological recording windows. This design isolates the three manipulated features (time of day, lighting characteristics, and trees). It enables estimating how each affects stress and related perceptions in this specific public space, in line with the research questions in Chapter 1.

Benefits of virtual reality. Virtual reality (VR) with an immersive head-mounted display is used as a tool for evaluating urban design and public spaces. VR environments allow site-specific simulations of complex or high-risk situations while preserving the controllability and repeatability of lab conditions (Feng et al., 2021). A systematic review of data-collection methods for studying pedestrian behavior supports the use of VR in experimental research on walking (Feng et al., 2021). Recent work further shows that VR can offer high ecological validity alongside experimental control that is difficult to achieve in real-world settings (Argota Sanchez-Vaquerizo et al., 2023), can capture citizens' perceptions of stress, safety, comfort, and place quality under systematically varied environmental conditions (Argota Sanchez-Vaquerizo et al., 2023; Bar-Ad et al., 2025; Ghanbari et al., 2024; White et al., 2023), and is suitable for evaluating lighting interventions and subjective responses in public space (Scorpio et al., 2020). In this study, the VR environment enables within-participant comparisons across all five scenarios while keeping all other attributes of the plaza constant, which would not be feasible in the actual station area.

3.2 Site Context and Analysis

The study focuses on the Bullewijk metro station plaza, a critical urban node undergoing rapid transformation. This section analyzes the site using quantitative data (statistics from CBS and Allecijfer), reports, and qualitative site observations (GIS, photo/video documentation and observations). This frames the challenges of safety perception and spatial design, supports the construction of the virtual environment, and identifies current conditions, which help the discussion.

3.2.1 Location within Amsterdam

Amsterdam's public transport is complemented by an extensive and reliable cycling network and a multimodal transit system that includes trains, trams, buses, and two main metro lines. Bullewijk metro station is located on the east corridor of the metro and is served by lines 50 and 54 (green and yellow), with an elevated track (GVB 2020). While cycling to work is common, the metro is a necessity because it is far from the city: 9 kilometers from the central station and 7 kilometers from Zuidas.

The station is situated in Amsterdam Zuidoost (The Bijlmer). This district was rapidly built in the 1970s with high-rise buildings over numerous open spaces, high diversity, and mixed incomes relative to the city average (CBS report Nederland). Municipality reports indicate lower perceived safety, particularly in the evenings (Buurtfocus, Nidi). Significant socio-economic contrasts and physical transformation characterize the area surrounding the plaza. Within a 500 m catchment mapped in Figure 3.1, on the west lies Amstel III A/B Noord, dominated by office and commercial buildings, some of which are mixed with residential, most of which are high-rises. A complex called "The Ensemble" is under construction, measuring 120 meters in height, and is located within the VR environment. The east side is primarily residential, with some student housing, and a few stores and small businesses located on the ground level.

Table 3.1 (Allecijfer.nl, based on CBS data, 2024) shows the differences: statistics for the adjacent neighborhoods are displayed. The eastern neighborhoods have incomes below the city average and a higher share of residents with non-European origins, whereas the western neighborhoods have above-average incomes. The older residential areas of Hoptille and Hakfort (eastern) are low-income and predominantly originate outside the Netherlands, with a crime rate focused on residential targets (stolen bikes/mopeds). Conversely, the transforming Amstel III A/B Noord, the plaza's western side, is becoming a high-income area (average annual income: €49,500) due to the construction of new residential buildings (indicating a growth from 85 residents in 2022) and its dominant commercial function.

Table 3.1: Statistics of the adjacent neighborhoods, compared to Amsterdam's average

Indicator	**Amstel III A/B Noord (western)**	**Hoptille/Hakfort (eastern)**	**Amsterdam (avg.)**
Population	1,150 residents	4,050 residents	≈933,000
Avg. Annual Income, per person	€49,500	€19,300	€38,600
% Residents Originating in NL	14%	10.90%	40.30%
Total Registered Crime Rate (per 1,000 residents)	690	70	90

Primary Crime Type	Theft of/from Motors (323 total, likely of commuting employees)	Stolen Moped/Bikes (Half of all crimes)	Thefts (General)
Key Spatial/Social Insight	High crime rates are attributed to the mono-functional use of offices and a lack of "eyes on the street" for surveillance. Attracts high-income commuters.	The area is characterized by low income and a high density of residents with non-European origins. Crime primarily affects residents (mopeds/ bikes).	Serves as the high-income, diverse core against which the Zuidoost district compares poorly.

3.2.2 Metro station vicinity

Bullewijk metro station has two entrances (Figure 3.1): North and South, both having plazas in East and West. A station plaza is the paved threshold of the station to the cross of the bike lanes. The entrances are under a viaduct of the metro and intercity train. The train tracks are elevated and serve as a significant physical barrier, dividing the eastern Amstel III A/B Noord office district from the western Hoptille and Hakfort residential neighborhoods. Within the 500-meter study radius, access between these zones relies on three passages, which are only within the station (at the north, south, and a road in between).

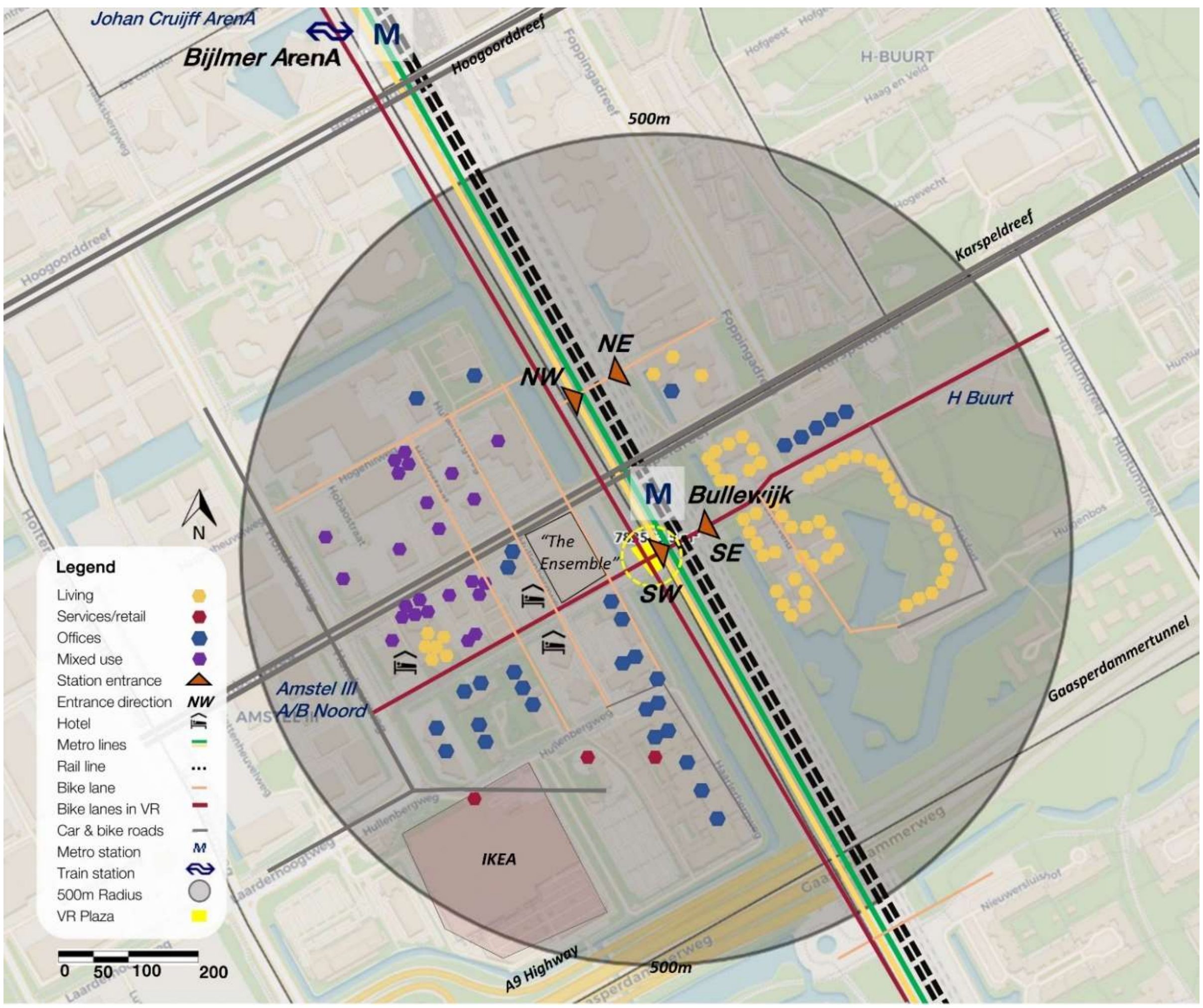

Figure 3.1: Aerial view of the Bullewijk metro station and its surroundings. The studied location is the southwestern (SW) entrance, depicted in the picture with a circle.

This channeling concentrates pedestrian and cyclist flow, and places pressure on the station plazas as mandatory crossing points. Furthermore, the western office-oriented district exacerbates the problem: large, empty spaces after work hours that famously lack "eyes on the street," which removes natural surveillance (Jacobs, 1969), contributing to a threatening atmosphere and dulling after-work hours. When the environment is empty, the few people who remain are often perceived as those engaged in activities that cause fear, making the presence of any person feel threatening. When underpasses are seen as a sudden change, a 'hard cut', this is quickly perceived as unsafe. The 'hard cut' causes unpredictability (Gehl, 2013; Rossetti et al., 2019; Schönfeld & Bertolini, 2017).

3.2.3 Detailed description

This study focuses on the south-western plaza (SW, Figure 3.2). In the VR experiment, the specific plaza is studied and recreated in detail. The plaza is defined by a large, continuous hardscape surface comprising asphalt, concrete, and pavement patches. Poor care cues, such as visible graffiti on the walls and litter, characterize the public space. The plaza has no amenities: no fixed seating, kiosks, restrooms, or vending services. There are no trees or canopy shade along the main pedestrian desire lines. The only cover is the bridge deck itself. From the intersection of the bike lanes, long, clear sightlines are present in all four directions. Further enhancing the unpleasant environment is the loud, periodic noise produced by trains passing every few minutes in the evening under the deck. At night, as shown in Figure 3.3, the lighting system includes a tall lighting pole that creates glare for approaching pedestrians, while the ground plane remains dark under the deck. The light is intense, cool, and unevenly distributed.

Figure 3.2: The Bullewijk metro station plaza, view towards the west. The entrance to the station is situated under the bridge.

Figure 3.3: The Plaza at night. The plaza itself is not lit, creating a dark spot on the ground.

The Bullewijk SW plaza is a clean test case for isolating the effects of designs on pedestrians. The Bullewijk station plaza remains minimalist**.** Understanding how to enhance safety in this plaza is crucial, given the station's physical layout, the mandatory passage through it, the reliance of locals and commuters on it, and the extreme socio-economic contrasts between adjacent neighborhoods.

3.3 Virtual Environment

As this experiment aims to test design tweaks in the site-specific environment of Bullewijk, Amsterdam, the environment needs to be precise, realistic, and convincing. Therefore, the scene was recorded with the greatest detail and precision, which was then transformed into a virtual environment. This section details the process of building a replica of the Bullewijk metro station in Amsterdam, into virtual reality.

3.3.1 Construction of the environment in Unreal Engine

In order to resemble reality, detailing is key (Feng et al., 2021). Following the site analysis, the location was replicated through photography and videos taken in the plaza at different times of the day, from different angles, using both stills and 360-degree videos, with sound recording.

Engine and setup. The scene was authored in Unreal Engine 5.5.3. World units were left at the default (1 Unreal Unit = 1 cm). The project uses Lumen for dynamic global illumination and reflections, and Sky Atmosphere, Skylight, and Directional Light stacks for physically-based skies, both during the day and at night. A global Post-Process Volume controlled exposure and grading. Motion blur was disabled.

Capture and Scale. Geometry and sightlines were reconstructed from site photos, 360-degree video, and tape records that were done by the author at the Bullewijk metro plaza in Amsterdam. Key dimensions, such as the train bridge, hills, and walls for the battery, the height of objects, and buildings, were measured and estimated using imagery and Google Earth and Google Maps measurement tools. The player's (participant view) height was set to 170 cm. The main Line-of-sight corridor, under the train bridge, was especially detailed. The scaled masses were then built using Rhinoceros to adapt to the Unreal Unit scale and then placed in the location using a Google Earth scale satellite patch.

Geometry and Assets. The scene combines some pre-made assets and some site-specific and unique objects and meshes, created for this experiment. Some street furniture, cycle path attributes, distant buildings, and pavements were used from the TU Delft XR NewMedia lab, and others, such as graffiti, wear, lightbulbs, foliage, trees and furniture. In addition, some objects were created and adjusted specifically for this scene, including streetlights, site-specific buildings, pavements, and gravel. Trees and foliage are assets from libraries, and those used are moving slightly with gusts to look more realistic.

Materials. Materials follow physically-based conventions with Base Color, Normal, Roughness, Metallic and other inputs in the Unreal Engine environment. Surfaces were tuned to match site photos: asphalt and concrete to metal railings, ceramic tiles, cycling paths, building facades, and glass windows.

Sun and Sky. A movable Directional Light (sun) and Skylight (captured from Sky Atmosphere) were set, according to Unreal's Sun Position controls to match typical Daytime (afternoon shade) and nighttime windows. During the nighttime, the sky is lit with minor moonlight, but stars are in sight. Lumen GI and Reflections were enabled, and Exponential Height Fog with Volumetric Fog added light haze and realism.

Artifact Light. During night scenarios, street lights and under-bridge luminaires were Moveable to reproduce real photometric spreads, with spot light and point lights, customized Intensity and adjusted Colors to match the wanted Kelvins: "white" cool lighting, the as-is baseline scenario, had ~5500 K, and the "yellow" enhanced light had all lightings, including those under the bridge and further streetlights changed to a warmer ~2700 K with a less intense Lumens. The sun lights were adjusted with Lux. The Unreal Engine units are similar but not necessarily identical to real values, so the most realistic values were chosen. Furthermore, Lumen handles most glossy reflections, but local Sphere Reflection Captures help stabilize highlights on large glass façades.

Figure 3.4: site as-is in reality and in virtual reality: top is reality and bottom is VR; left is night and right is day.

Audio and Train Event. Ambiance and events were implemented with Sound Cures. Day tracks featured diffuse urban activity and distant traffic. Birdsong was added in Scenario B near the trees. A train pass-by (from 55s to 72s) was synchronized to visuals with Sequencer, including Doppler enabled on the audio component and a moving attenuation cone to simulate approach and recede. The train event appeared similarly in all scenarios and was lit at night. All files were stereo 48kHz, normalized for consistent perceived loudness. The audio during the night was dramatically quieter, and in both day and night, some distant chatter could be heard, but not a distinct language, meaning, or number of persons in sight.

Performance. Capping reflection quality was set to the default “High”; post-process volume used neural color grading; vignette, grain, and lens flare were minimized for comfort and smooth operation of the scenarios. The timestamps script automatically recorded the entrance and exit times for each scenario.

3.3.2 Scenario design

The experiment consists of five scenarios and a familiarization phase. The familiarization phase, the first encounter of participants with the VR, was the respawn studio of UE. Participants were asked to walk through the room, understand the scales and boundaries, find the boundaries as they approached the wall, detect mishaps that sometimes occur, and calibrate the volume by hearing the fireplace sound through the VR headset (set to 7/12 bars). Table 3.2 indicates the scenarios list with its attributes, and Figure 3.5 to Figure 3.9 shows sights from scenarios A to E.

Table 3.2: scenarios list

Scenario	Time	Trees	Lighting	Extra Audio	Title
A	Day	-	Daylight	-	Day as-is, baseline
B	Day	Enhanced	Daylight	Birds, located in the trees	Day, more trees
C	Night	-	Cool, White, ~5500 K	-	Night as-is, baseline
D	Night	-	Warm, Yellow, enhanced and distributed more even, ~2700 K	-	Night, enhanced lighting
E	Night	Enhanced	Warm, Yellow, enhanced and distributed more even, ~2700 K	-	Night, enhanced lighting, more trees

Figure 3.5: Scenario A: Day as-is. Looking forward from the respawn position, towards the metro station entry.

Figure 3.6: Scenario B, Day enhanced trees. Three trees to the left, one large tree to the right, and two on the back were added.

Figure 3.7: Scenario C, night as-is. The plaza is identified by a dark spot. The large light pole in the center lit the periphery.

Figure 3.8: Scenario D, night with enhanced lights. Lighting poles were allocated in the dark spot of the plaza to ensure fair distribution of light, leaving no shady spots.

Figure 3.9: Scenario E, night with enhanced lights and enhanced trees. Lights were distributed to ensure minimal shadow of trees.

Participant Experience. In all scenarios, participants could freely walk and change viewpoints within the 2 x 2 meters box. By physically walking, the virtual reality perspective changes. Moving the head and body naturally, changes the sight. There were no interactive tasks, no teleportation using the sticks, and attention was not directed anywhere. Participants respawn in all scenarios with their head facing the entrance of the metro station. The headset was connected to the computer via a USB wire.

Figure 3.10| Participant experience. left: the respawn box: participants can walk physically within these borders. Right: a satellite view of the VR environment – the lines of sights that participants can see and engage in.

Scenario Length and Events. Each scenario takes 90 seconds, from the moment of spawning in the environment until a closing message pops up on the screen. There are no pedestrians or any other characters appearing in either scenario, besides a passing train, which appears in all scenarios starting from second 55. During night scenarios, the train is lit.

Enhanced Trees. Three 4 meters trees were placed on the northern (left) side of the plaza, one 6 meters tree was placed on the southern (right) smaller side of the plaza, and three other trees on the grassland just outside the plaza. The trees were set according to Gehl's theory of urban design and the literature (Gehl, 2013), to both be visible and increase benefits of trees but also not block the main sight of passing travelers, to avoid unpleasant hiding or loitering people, and let passing by people see who comes and goes out of the metro. The three trees are creating a line that allows activity, leads direction, but does not leave out of sight dark spots during nighttime (Gehl, 2013).

Enhanced Lighting. The lighting in the night, as-is baseline scenario, is characterized by a large light pole with five strong projecting lights, using a cold-white with 5500 kelvins. The five large bulbs were distributed evenly 360 along the pole, leaving a distinct large dark spot under the pole itself (*figure 3.4*). Therefore, following Rabaza et al. (2013) and Gehl's theory, the lighting conditions in the enhanced scenarios were distributed more evenly, two-sided staggered and provided with a warmer yellow 2700 Kelvin lights. The design of the lighting was kept simple, with simply additional "Amsterdam Style" precast street light poles. This decision was made to avoid over-design, and keep the lighting condition as isolated as possible. Possible redesign of the lighting conditions in the plaza should take more care of the appearance of the design, or introduce creativity in lighting.

Figure 3.11: Up, left and clockwise: enhanced trees main plaza; enhanced tree small plaza; metro sign and as-is lighting poles; lighting pole during day scenarios; lighting poles during night as-is (cool light).

3.4 Experiment apparatus and setting

3.4.1 Hardware

In this experiment, participants were immersed in the virtual environment with a Meta Quest 3 head-mount display, connected via a wired connection (USB) to a desktop computer (Intel®CoreTM i7-11700 11th generation [8 Cores, 16 MB Cache, 2.5 GHz up to 4.9 GHz], NVIDIA®GeForce®RTX 3070TM 8 GB GDDR6; 32 GB M.2-PCIe-SSD with 1 TB + SATA with 1 TB, 7200 1/min). To gain an understanding of how public space design influences stress levels of participants, participants' physiological responses were captured with an Empatica E4 wristband connected to a smartphone via Bluetooth. It synchronized and stored the data anonymously in the CareLab Portal Cloud. Earlier in the process, to record the public space's details and ambiance, 360-degree video and still photos were captured with an Insta360 camera.

3.4.2 Software

The VR environment was developed, built, and rendered in Unreal Engine 5.5.3. Some elements were built and modeled in Rhino; others were sourced from the TU Delft XR lab library, and others were obtained from the Unreal Engine Asset Store. A few designs were created using scripts written in C# (to time the train's movement and capture timestamps of each session). The Empatica Health Monitoring Platform was used on the CareLab app, and the AWS Python portal was used to access raw data.

For data analysis, custom scripts were developed for this study to preprocess, clean, align, and extract features from physiological data using Python, launched via Visual Studio. For extracting, preprocessing, and performing model calculations on the physiological data, R was used via RStudio. Questionnaire data were collected on Microsoft Forms. Microsoft Excel was used to analyze and read data. The Calendly app was used to schedule and manage timeslots for the experiments, and Adobe Illustrator was used to design the poster. Acknowledgement of use of AI and technical support appears in *Appendix J*.

3.4.3 Room configuration

The experiment took place in two similar locations: at the AMS institute, in the “Cairo” room, and the TU Delft campus library “Studio”. Out of 47 sessions, 38 were conducted in Amsterdam and 8 in Delft. The room in both locations had identical characteristics for the participants: both had an approximate 2*2 square to walk in, next to a table with a screen and a chair to complete the questionnaires on. Both rooms were quiet and had a controlled temperature of 24-25 degrees Celsius. Both rooms were small, with 10 square meters in total, and had a neutral odor. The room in Delft had a darker environment because it is set to be a recording studio, while the room in Amsterdam had a window that was covered. However, since the experiment is done in a remote virtual reality setting, the lighting of the physical room had minimal to no influence on the results, and therefore was ignored.

Figure 3.12: left: Cairo room in Amsterdam; center: participant filling in a questionnaire during a session; right: Studio in Delft.

3.4.4 Recruitment procedure

The recruitment of participants was done via two main channels, in accordance to the two locations offered for the experiment: the majority of participants were recruited in Amsterdam, mainly from the AMS institute community (of which the author is affiliated with), with 83% (*n=39*) of the participants conducting the study in Amsterdam, and the rest in Delft.

All participants were volunteers and did not receive any compensation. Participants were recruited via a Calendly link, which allows them to schedule themselves into available time slots throughout the week. Posters were displayed in the AMS Institute in Amsterdam, along with those in the adjacent CODAM coding academy. Invitation posters were hung in several locations, both in Delft and Amsterdam, and were shared via WhatsApp student groups and faculty email distribution lists.

3.4.5 Experiment procedure

The session lasted approximately 30 minutes and followed this six-stage sequence:

1. Arrival and briefing. Participants read the briefing paper, provided informed consent, and were fitted with the Empatica E4 wristband on the non-dominant wrist. For full informed consent and briefing paper, see Appendix B.

2. Baseline questionnaires. Demographics, PSS-4, and the baseline SSSQ questionnaire were completed before exposure. See Appendix D.

3. Familiarization. Participants walked the room and learned the safety boundaries in VR. A neutral environment was presented to encourage natural movement and to check audio-visual quality. Connection, wires and smoothness was checked visually with participants. Sometimes, this stage was restarted to ensure similar conditions to all participants.

4. Five scenario windows and in-situ ratings. Each participant received a randomized order of the scenarios. Each scenario takes 90 seconds. If the visual or audio quality degraded, the window was aborted within a few seconds and reloaded. Participants were instructed to be aware of their environment, the length of the window, and that they could walk within their box. After each exposure, the SSSQ and the three single-item appraisals were completed.

5. Post-experiment questionnaires. SSQ, iPQ, the post-process "Use-Post" item, and a feedback open question were administered. At the Use-Post item, participants were reminded of their order of scenarios.

6. Debrief. The Empatica E4 wristband was removed and charged, and participants were offered a debrief and the option to receive study results.

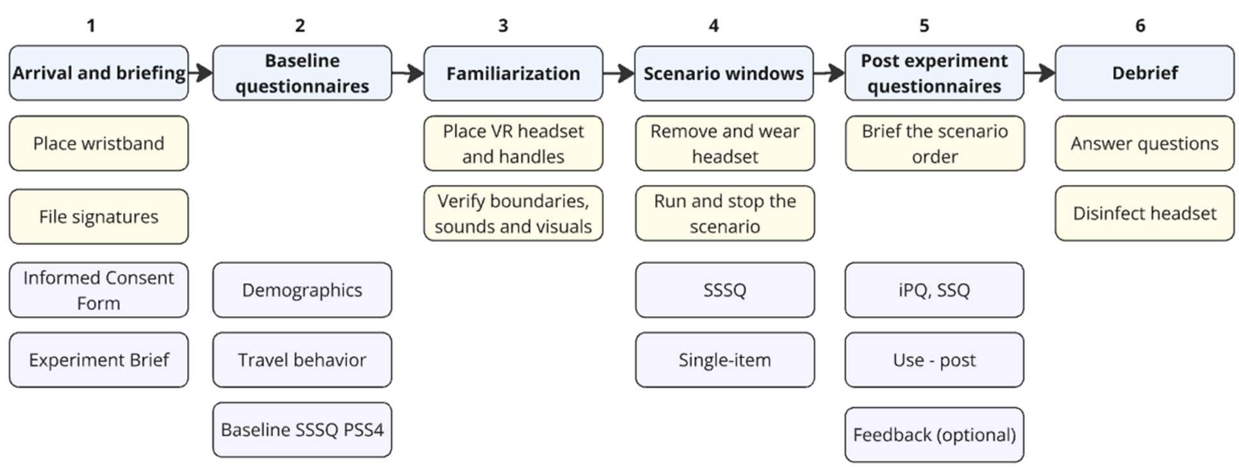

Figure 3.13 – Experiment procedure. Stages 1-6 (blue), Researcher's tasks (yellow), participants' questionnaires (violet).

3.5 Data Collection

Data were categorized into physiological and psychological channels. Physiological signals of Electrodermal Activity (EDA), Blood Volume Pulse (BVP), skin temperature, and 3-axis Accelerometer (ACC) were continuously collected using an Empatica E4 wristband worn on the non-dominant wrist. This continuous recording began with the initial familiarization phase and lasted until the end of the final scenario. Wear quality was verified on the E4 application before familiarization, and device clocks were synchronized for later alignment of data epochs using a custom Python script that matched Unix UTC timestamps from the Empatica export to scenario logs. Primary physiological outcomes derived from this stream were tonic Skin Conductance Level (SCL, µS) and Heart Rate (HR, bpm), with skin temperature as a covariate to account for SCL thermal drift. Conversely, psychological data were collected via self-reported questionnaires administered via Microsoft Forms, both pre- and post-exposure. See Appendices C, D, E, F, and G for the full questionnaires.

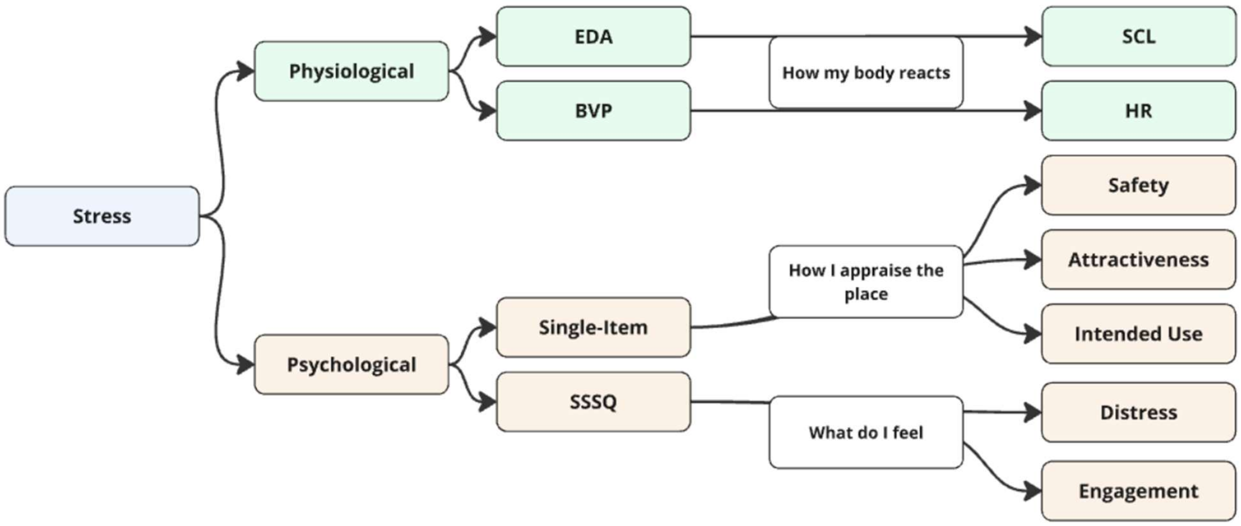

Figure 3.14: operationalization of perceived stress. Orange is the psychological channel, collected with questionnaires, and green is the physiological channel, collected with sensors (Empatica E4)

3.5.1 Psychological measurements

Short Stress State Questionnaire (SSSQ). Helton's 10-item version questionnaire was used, with the Engagement and Distress subscales. Items were rated on a 1-5 Likert scale (Not at all to Extremely). Subscale scores were computed as the mean of their items, with higher scores indicating more of the construct. The SSSQ was asked six times during the experiment: once before exposure, and then after each scenario. Rationale: Distress captures unpleasant arousal and overload, while Engagement indexes motivated attention. Together, they complement autonomic measures (Argota Sanchez-Vaquerizo et al., 2023; Helton, 2004).

The engagement and distress subscales are the primary outcomes, along with the three single-item appraisals per scene (Safety, Attractiveness, Use) and the Use-Post. The PSS-4 assesses baseline stress disposition, which is a descriptive outcome. The SSSQ-10 has two components: Engagement (questions 2, 5) and Distress (questions 1, 3, 4, 6, 7, 8, 9, 10). Baseline was set to the pre-exposure SSSQ. The appraisals per scene are single-item charts of "Safe", "Attractive", "Use", and "Use-Post" which were summed according to the key of 1-5. Furthermore, the design factors of Night (scenes C, D, E), Yellow (scenes D, E) and Trees (scenes B, E) were labeled.

Three single-item appraisals. After each exposure, and next to the SSSQ questionnaire, participants were asked as follows: "Please select how relevant these statements are to your experience in the previous scenario". They rated each statement on a 1-5 Likert scale (Not at all to Extremely). The statements were: "I found the previous scenario attractive as a pedestrian" ("Attractive"); "I would be willing to use the metro given the previous scenario ("Use"); and ("I felt safe in the previous scenario") ("Safe"). Higher scores indicate a positive perspective on the item. These appraisals were asked after each of the five exposures. Rationale: three items target perceived safety, attractiveness, and behavioral intent relevant to the scenario, and complement the main SSSQ (Sanchez et al., 2024).

Post-process Likert chart. After participants were shown all scenarios, they were informed of their particular order of exposure and then asked, "Given these scenarios, how likely are you to use the metro?" This question was rated on a 1-5 Likert scale (Never – Very likely), where a higher score indicates a positive perception of the scenario. All five scenarios were asked. Rationale: This outcome is labeled "Use-Post" and serves as a post-hoc preference summary.

Perceived Stress Scale, 4 items (PSS-4). Participants completed the questionnaire (Cohen et al., 1983) at baseline. Items were scored on a 1-4 Likert frequency scale. Higher scores indicate greater perceived stress. Rationale: provides a baseline stress benchmark.

3.5.2 Physiological measurement

Electrodermal Activity (EDA). Sampled at 4 Hz**,** measured with µS. Decomposed by the Empatica device and aggregated per window into tonic and phasic components. Tonic level is the slowly varying skin conductance level (SCL). Phasic activity (SCR) is the rapid peaks due to discrete sympathetic events. Lower EDA activity indicates a calmer individual. Phasic detections were unreliable after cleaning, given the brief windows and heavy movements. Therefore, only tonic SCL means were retained.

Blood Volume Pulse (BVP). Sampled at 64 Hz via photoplethysmography sensor (PPG). The PPG sensor measures light absorption in the skin to infer blood volume changes with each heartbeat. From the BVP waveform, the device provided inter-beat intervals (IBI, measured in milliseconds (ms)) – the time between successive pulse peaks. In ECG, the beat-to-beat interval is called RR, and with PPG, the IBI is used equivalently. After artifacts and ectopic beats were removed by quality control (QC), this series is referred to as normal-to-normal (NN). This interval series is the basis for computing both heart rate (HR, measured in beats per minute (bpm)), which is 60,000 / IBI (ms), and the heart rate variability (HRV), which indicates the Root Mean Square of Successive Differences (RMSSD, measured in ms) and Standard Deviation of NN intervals (SDNN, measured in ms).

In short-term recordings, such as the 90-second window, higher HRV indicates a stronger parasympathetic activation, which means a calmer or less stressed participant. HR means were used in the analysis, but the HRV metrics were discarded due to high correction rates. HRV metrics are extremely sensitive to corrections, gaps, and short window lengths, making them unreliable in this case.

Accelerometer (ACC) and Skin Temperature (TEMP). Temperature sampled at 4 Hz and used as a covariate. ACC sampled 32 Hz or a 3-axis, and summarized to magnitude per scene for descriptive confirmation of movement. All participants had a 99-100% "in motion" outcome. ACC was explored for motion masking but not used as an artifact mask in final EDA outcomes.

3.5.3 Other questionnaires

Simulator Sickness Questionnaire (SSQ). Administered post-session to quantify discomfort, sickness and pain (Kennedy et al., 1993).

igroup Presence Questionnaire (iPQ). Quantifies the sense of presence as the subjective sense of being in a VR environment. This scale measures spatial presence, involvement and experienced realism (Schubert, 2003).

Demographics, travel habits, and familiarity with VR. Participants reported gender, age, current postal code, upbringing country, neighborhood context, prior VR exposure, academic background, driving license, car ownership, and weekly kilometers, usual transport modes, public transit frequency, familiarity with Bullewijk station, metro usage frequency, and any prior severely unsafe metro experiences. The full instrument appears in Appendix D.

3.6 Ethics and data management

To initiate research involving human participation, as per this experiment, an ethical approval had to be submitted (in accordance with the 1964 Helsinki Declaration and the Nuremberg Principles). The research protocol was reviewed by the TU Delft Human Research Ethics Committee (HREC) and was approved with the reference number #5008. The full approval with data management plan is in *Appendix H*

All participants volunteered to participate in the experiment, one by one, and did not receive any reward or compensation for their participation. Each participant was assigned a participant ID, and apart from the demographics questionnaire, the gender, age, country of origin, participant's data, and responses could not be traced back to their personal information. Furthermore, the study follows the strict guidelines of the TU Delft HREC to ensure confidentiality. Access to the recorded data is restricted to the responsible researcher and the supervisor only.

The first part of the experiment begins with reading and signing the informed consent form (*Appendix A*), along with an information sheet regarding the study's purpose, procedure, and the participant's right to withdraw from the experiment at any point, without providing a reason.

Following the TU Delft HREC guidelines, data were stored on a secure server. They were accessible only to the supervisor and the responsible researcher to protect the confidentiality of the participants. Raw physical data were collected and stored on Empatica's secure server, accessible only to the supervisor and the responsible researcher via a password. The anonymized data will be archived for 10 years on encrypted servers. The data collected in this study will not be shared with external parties or researchers, and the findings from this study will be solely used for academic purposes.

4 Measures and Data Analysis

This chapter reports on the participants, the data collection methods, both physiological (objective) and psychological (subjective), the data processing and quality control procedures, the statistical analyses, including variables and model structure, and the statistical tests and models used. Stress is conceptualized as described in Section 2.3. For the data cleaning protocol, see Appendix A.

4.1 Participants

The sample size consisted of 47 participants, out of which 41 were included in both physiological and psychological channels. The remaining 6 had technical issues with either the Empatica device's connection or timestamp capturing, and were therefore discarded from the physiological data before quality checks. See Appendix A. During the analysis phase, 3 of the 41 participants were excluded because they failed QC across all windows. Table 4.1 describes the data used from participants in each channel. Shapiro-Wilk indicated deviation from normality ($p<0.001$).

Table 4.1: Demographics. In the physiological analysis, out of the 41 participants which were included without technical issues, 3 were removed due to unreliable sensors outcomes, resulting in 38 participants which eventually contributed to the outcomes.

Channel	Sample size	Age Mean (years)	Age SD (years)	Age Median [IQR]	Age range (years)	Gender
Psychological	47/47	29.23	8.51	27 [24-32]	20-64	Male 26 (55.3%), Female 20 (42.6%), Prefer not to say 1 (2.1%)
Physiological	41(38)/47	28.11	6.64	26 [24-30]	20-55	Male 22 (55%), Female 18 (45%)

4.2 Psychological Measures

4.2.1 Variables, scoring, and model structure

Dependent variables (DV). Continuous variables: SSSQ-10. Two subscales: Engagement (Q 2,5; mean of items). Distress (Q 1,3,4,6,7,8,9,10; mean). Response scale 1–5 ("Not at all" – "Extremely"). Ordinal variables: Appraisals per scene. single-item **Safety, Attractiveness, Use**, and **Use-Post** (post-exposure likelihood to use the metro) (1–5).

Independent variables (IV). Scene-level predictors: Design factors from scenes A-E. Night (C, D, E), Yellow (D, E), Trees (B, E).

Covariates. Participant-level: PSS-4 (centered PSS-4 total), Age (centered), Gender (factor). Scoring PSS-4: items 2 and 3 were reverse-scored, and the total score was summed. Higher totals indicate greater perceived stress. Baseline covariates: SSSQ-10. The **baseline** SSSQ was collected before exposure.

Random effect. Each participant received a random intercept to model individual baseline differences across repeated measures.

4.2.2 Data checks

Internal consistency: Cronbach's alpha. Computed per subscale (Engagement (k=2, expected to be modest), Distress(k=8)). This check assesses the coherence of items within a subscale. Alphas between 0.7 and 0.9 are often considered "good".

Distribution screens: Shapiro-Wilk and Q-Q plots. Tests normality per scene. Small p-values indicate non-normality ($p<0.5$). Q-Q plots visualize normality: if points lie close to the 45-degree line, normality is plausible.

Variance homogeneity: Levene's test. For the continuous data. Checks equal variances from the median-centered across scenes for continuous outcomes, under a one-way ANOVA on deviations across A-E scenes. A non-significant p-value ($p \geq .05$) indicates that there is no evidence of heteroscedasticity.

Outlier detection: Within-scene z. Flags extreme values within each subscale x scene. $|z|>3$ is marked "extreme", and in sensitivity checks will be omitted. **Cook's D** was screened and reported in Appendix A, but no exclusions were made based on this test (missingness was minimal).

Ordinal modeling assumption: Proportional-odds (PO) per IV. CLMM assumes each IV shifts the latent log-odds by the same amount across all cut points (parallel slopes). PO reports a single odd ratio (OR) per IV. PO was tested using likelihood-ratio (LR) χ^2 comparisons between a PO model (one slope per predictor across thresholds) and an augmented nominal-effect model (predictor allowed to vary by category). LR χ^2 tests were nonsignificant, indicating no difference in effect by threshold, supporting the PO.

4.3 Physiological Measures

Sensors collected skin conductance level (SCL, µS) and heart rate (HR, bpm). For each participant and each scene window (A-E), mean SCL and mean HR were computed over the full 90 seconds. Skin temperature was included as a participant-centered covariate to control tonic drift (in SCL), and ACC was used to confirm that movement occurred. ACC was not used for EDA masking.

4.3.1 Variables and model structure

Each DV was summarized as mean ± SD by scenario and as participant-level medians to show within-person central tendency. QC yields were tabulated by channel and scenario to contextualize inference.

Dependent variables (DV). Skin Conductance Level (SCL) (mean, µS) and Heart Rate (HR) (mean, bpm).

Independent variables (IV). Scene-level predictors: Design factors from scenes A-E. Night (C, D, E), Yellow (D, E), Trees (B, E).

Order and trend. Order (1 to 6), First (0 or 1, first post-baseline scene, often a "jump" due to settling in), and a linear Trend (0 to 5) were tracked for habituation checks. Because SCL and HR are time and arousal-sensitive, these are relevant here (but not for the psychological outcomes).

Covariates. Gender and age were explored in model variants for SRQ4. For the SCL, the temp-centered part was the only covariate (to identify within-person thermal drift).

Random effects. Repeated measures were handled with a participant random intercept in linear mixed models (LMM).

4.3.2 Quality control rules

A window passes quality control (QC) if it passes either EDA or HR.

EDA cleaning and inclusion. Because participants were instructed to move, the ACC-based motion mask would blank most seconds. As described in the pipeline, scenes were retained if the percentage of EDA samples missing or masked after cleaning, divided by the total expected EDA samples in the scene was at most 10%. Otherwise, they were marked EDA - QC fail, and excluded.

HR signals. HR was summarized as the mean beats per minute (bpm) per scene. HRV was pre-specified to require at least 60 seconds of contiguous clean RR and at most 5 percent corrected beats, with any gap longer than 3 seconds disqualifying the HRV window. HRV was not reported (RR median = 38.7%).

4.4 Statistical Analysis Plan

The analysis proceeds in three steps. First, scene-level descriptives and data quality checks are reported. Second, planned contrasts address STQ1-SRQ3 for both channels. Third, mixed models estimate adjusted scene effects and test moderation by gender and age for SRQ4, using repeated measures.

4.4.1 Descriptive analysis for psychological data

For scenes A to E, psychological continuous variables (Engagement and Distress) were summarized as mean ± SD and median [IQR]. Ordinal items (Safety, Attractiveness, Use, Use Post) were summarized as median [IQR]. Internal consistency (Cronbach's alpha), Shapiro Wilk tests, Q-Q plots, Levene's tests, and within scene z scores (outliers with $|z|>3$) were reported for the SSSQ subscales.

4.4.2 Planned Contrasts

Psychological planned contrasts: Wilcoxon test. A planned contrasts test was conducted with Benjamini-Hochberg False Discovery Rate (BH FDR) control for each DV, following the contrasts identified in the introduction, per SRQ. Effect sizes were reported as the median within-person Δ with 95% bootstrap confidence interval (CI) for all DVs, and the BH-FDR adjusted p value. The within-participant delta was

$$\Delta_i = y_{i,\text{left}} - y_{i,\text{right}}$$

for participant i on a given contrast. As a robustness screen, Friedman's omnibus test was applied across the five scenes to detect any systematic within-participant ordering. Kendall's W was reported as an effect size, complementary to the contrasts (indicating the strength of the pairwise effect).

Physiological planned contrasts: Wilcoxon test. For SCL and HR, the same Wilcoxon signed-rank contrasts and BH FDR procedure were used, following Table 4.2 and the SRQs. Median within-participant deltas and 95 percent bootstrap CIs were reported, and Friedman's omnibus tests with Kendall's W were used to screen for systematic ordering of scenes. These tests were applied separately to SCL and HR.

Table 4.2 summarizes the mapping between contrasts and SRQs.

Table 4.2: Planned contrasts, answering SRQs

SRQ	Question Focus	Contrasts	Rationale
SRQ1	Day vs Night	A | C	As-is Day & Night
		B | E	Enhanced trees scenarios
SRQ2	Lighting at Night	C | D	Night without trees
		C | E	Night: As-is vs Trees
SRQ3	Trees	A | B	Day scenarios: without vs with trees
		D | E	Night enhanced light scenarios, with vs without trees
SRQ4	Moderation by Gender and Age	Screen all of the above with the moderations of Gender and Age	

4.4.3 Linear Mixed Models

Mixed models were used to estimate covariate-adjusted scene effects while accounting for repeated measures within participants and between-person differences. Continuous psychological outcomes (Engagement and Distress) and physiological outcomes (SCL and HR) were analyzed with linear mixed models (LMM). Ordinal appraisals (Safety, Attractiveness, Use, Use Post) were analyzed with cumulative link mixed models (CLMM) under the proportional odds framework.

Psychological Linear Mixed Model (LMM) for the continuous DVs. This model estimates scene effects while accounting for repeated measures and between-person differences. The model includes fixed effects on the IVs, adjustments for the baseline covariates, and adjustments for trait stress from the participant-level covariates (PSS-4). A random intercept was included for each participant (using best linear unbiased prediction BLUP). The model is used because it retains all available observations, adjust for measured differences, and account for the within-participant correlation structure. The model was

$$y_{ij} = \beta_0 + \beta_{(1\,\mathrm{Night_j})} + \beta_{(2\,\mathrm{Yellow_j})} + \beta_{(3\,\mathrm{Trees_j})} + \gamma\,\mathrm{Baseline_i} + \delta\,\mathrm{PSS4_i} + u_{0i} + \varepsilon_{ij}$$

Where y{*i,j*} is the Engagement or Distress (DV) score for participant {i} in scene {j}. The coefficients β1, β2, and β3 quantify the adjusted mean difference associated with Night, Warm light, and Trees (IV), expressed on a 1-5 scale. *Baseline{i}* is the participant's SSSQ0 (baseline covariate) subscale mean, and γ captures the association between the baseline level and the scene scores. *PSS4{i}* is the centered PSS-4 total, and δ captures the trait stress adjustment. The *u{0i}* is the participant-specific random intercept that captures stable individual differences across all scenes. *e{ij}* is the residual error for that observation. In conclusion, the SRQs are answered using the coefficient beta for IVs: a positive beta indicates an adjusted increase in the outcome in original units, and a negative beta indicates an adjusted decrease. Confidence intervals and p-values quantify precision and evidence.

Physiological Linear Mixed Model (LMM) for the continuous DVs. A linear mixed model was fitted to mean SCL and to mean HR to estimate the effects of the IVs while accounting for repeated measures. A participant random intercept was specified in all models (BLUP estimator). Skin temperature was included as a within-participant covariate for the SCL models by using the participant-centered scene temperature, to identify arousals originating from different temperatures. The one-line specification used was

$$yij = \beta 0 + \beta 1\, Nightj + \beta 2\, Yellowj + \beta 3\, Treesj + \zeta\, tempij + u0i + \varepsilon ij$$

where *y{ij}* is mean SCL or mean HR for participant i in scene j. The coefficients β1, β2, and β3 represent adjusted mean differences associated with the scene attributes. The coefficient ζ captures the within-participant association between skin temperature and tonic SCL. The term *u{0i}* is the participant random intercept and ε{ij} is the residual error.

Psychological Cumulative Link Mixed Models (CLMM) for the ordinal DVs. Ordinal outcomes were analyzed under the proportional-odds (PO) framework (the probability of rating at or below each category), while using one effect per predictor. Fixed effects were the IVs, PSS4 was included as a covariate. A participant random intercept was included so that each individual had a personal baseline level (using BLUP). This term was estimated as part of the mixed-effects fit and accounts for the correlation of repeated ratings within the same participant. The form is:

$$ln(1 - Pr(Yij \leq c)Pr(Yij \leq c)) = \theta c - (\beta 1Nightj + \beta 2Yellowj + \beta 3Treesj + \delta PSS4i + u0i),$$
$$c = 1, \dots ,4$$

Here θc are the ordered cutpoints that partition the latent response into the 5 Likert categories. The linear predictor inside the parentheses contains the fixed effects for scene attributes and trait stress, together with the participant random intercept $u0i$, which captures stable individual differences and the correlation across repeated ratings.

For a predictor with coefficient β, the odds ratio is exp(β). Values of exp(β)>1 indicate higher odds of giving a higher category when that attribute is present, holding other terms constant. Values below 1 indicate lower odds. The proportional odds assumption was assessed using likelihood ratio tests comparing the proportional odds model with augmented models that allowed category-specific deviations in predictors. No violations were detected, so a single odds ratio per predictor was reported. SRQ1–SRQ3 are answered using these odds ratios.

Moderation (for SRQ4). Gender and age were tested as moderators of the IV. For LMM and CLMM, LR χ^2 tests compared models with and without interactions. Furthermore, multiplicity control for moderation was applied: BH-FDR across interaction tests per outcome to decide which interactions to follow up. Where significant, simple effects were summarized as gender-specific OR multiplier per +10 years of age.

Sensitivity. Sensitivity analyses excluded observations with |z|>3 within-scene for continuous psychological outcomes and refit the models. Conclusions about scene effects were unchanged. For the physiological channel, alternative QC thresholds and alternative HR processing variants were explored; the reported specification reflects the stricter QC rules, which produced stable inferences.

5 Results

This chapter presents the results of this study across two channels of psychological and physiological outcomes, with each outcome reported separately by sub-research question (SRQ) and by gender and Age. The analysis uses descriptive statistics to describe the initial outcomes, then applies them to the two main models: planned contrasts and mixed models.

5.1 Sample description

Forty-seven participants took part. The sample consisted of 26 men, 20 women, and 1 individual of unknown gender, with childhood origins in 18 countries (43% from the Netherlands). Participants reported growing up mostly in compact, walkable areas (43%), followed by low-density suburbs (23%), rural/village (19%), and city centers (15%). Currently, 77% live in a compact, walkable neighborhood (13% city center; 11% low-density suburb). The academic background was mainly in Made (49%), with Built Environment (15%) and other fields (36%). While 87% hold a driving license, only 17% own a car (43% have access to a family/friend's car); 10% drive more than 50 km weekly. Public transport use is frequent: "very often" (34%), "often" (32%), "sometimes" (30%), and "seldom" (4%). No participant lives near Bullewijk; 57% had never been there. Using Amsterdam's metro is "sometimes" for 43% and "seldom" for 34%; 4% reported ever feeling unsafe in Amsterdam's metro (11% elsewhere). Prior VR exposure was limited (34% never, 40% seldom, 26% sometimes). Most of the participants were students, with 49% (*n=23* having a background in the MADE program (Metropolitan Analysis, Design and Engineering).

5.2 Psychological outcomes

Scene-level descriptives for Distress and Engagement (SSSQ) and the ordinal appraisals (Attractiveness, Safety, Use, Use-Post) are shown in Table 5.1. Across scenes, median appraisals clustered toward the favorable end for Safety, Use, and Use-Post, while Engagement and Distress varied more modestly. As shown in Figure 5.1, the night-cool scene (C) typically shows the least favorable pattern (higher Engagement/Distress; lower appraisals), while day scenes (A/B) and night warm scenes (D/E) are rated better. IQR values in brackets refer to the 25th and 75th percentiles.

Table 5.1: Descriptive statistics for psychological variables of scenes A-E.

Scene	Distress mean	Distress SD	Distress med. [IQR]	Engage. mean	Engage. SD	Engage. Med. [IQR]	Attractive med. [IQR]	Safe med. [IQR]	Use med. [IQR]	Use Post med. [IQR]
A	1.26	0.41	1.12 [1.00, 1.38]	2.29	0.75	2.00 [2.00, 2.88]	3.00 [3.00, 4.00]	4.00 [3.00, 4.00]	4.00 [4.00, 4.00]	4.00 [4.00, 5.00]
B	1.22	0.38	1.12 [1.00, 1.31]	2.31	0.7	2.00 [2.00, 3.00]	4.00 [3.00, 4.00]	4.00 [3.00, 4.00]	4.00 [3.00, 4.00]	5.00 [4.00, 5.00]
C	1.4	0.6	1.12 [1.00, 1.50]	2.87	0.85	3.00 [2.50, 3.50]	2.00 [1.00, 3.00]	2.00 [1.50, 3.00]	3.00 [2.00, 4.00]	3.00 [2.00, 4.00]
D	1.26	0.36	1.12 [1.00, 1.40]	2.49	0.82	2.50 [2.00, 3.00]	3.00 [3.00, 4.00]	3.00 [3.00, 4.00]	4.00 [3.00, 4.00]	4.00 [2.00, 4.00]

E	1.25	0.43	1.12 [1.00, 1.34]	2.46	0.9	2.50 [2.00, 3.00]	3.00 [2.00, 4.00]	3.00 [2.25, 4.00]	3.00 [3.00, 4.00]	4.00 [3.00, 4.00]

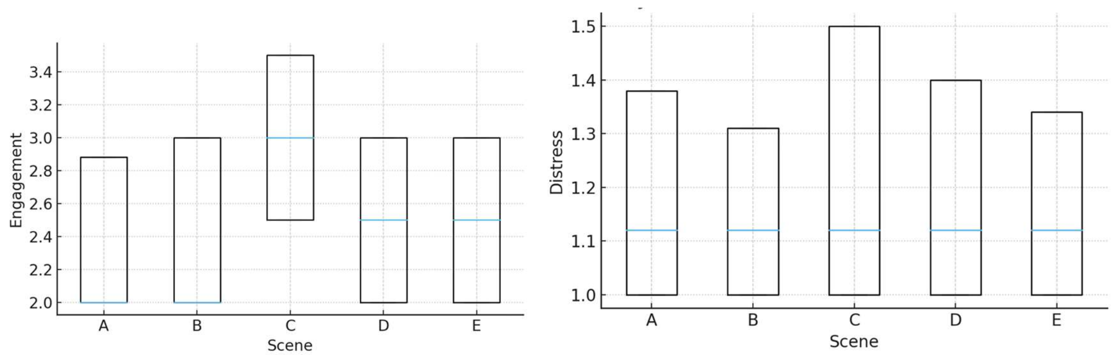

Figure 5.1: Boxplots for Engagement (left) and Distress (right) per scene A-E.

Internal consistency, as shown in Table 5.2, was acceptable for Distress (α≈.87, k=8) and modest for the 2-item Engagement subscale (as expected). Distribution checks, as shown in Table 1.3, suggested clear non-normality in most Distress and some Engagement cells.

Table 5.2: Internal consistency and outlier detection for Engagement and Distress, out of the SSSQ (k=10) questionnaire.

Subscale	k	alpha	Levene p	Total outliers
Engagement	**2**	**0.445**	**0.384**	**0**
Distress	**8**	**0.867**	**0.535**	**8**

Table 5.3: Shapiro-Wilk distribution analysis for Distress and Engagement variables.

Scene	Subscale	W	p	Scene	Subscale	W	p
A	Distress	0.654	<.001	A	Engagement	0.928	0.007
B		0.622	<.001	B		0.925	0.005
C		0.676	<.001	C		0.943	0.024
D		0.743	<.001	D		0.946	0.032
E		0.617	<.001	E		0.940	0.021

5.2.1 Planned contrasts

The pre-registered within-participant Wilcoxon signed-rank contrasts (Table 4.2) provide a picture of how time of day, lighting, and trees shaped scene appraisals. Results are summarized from the strongest to the weakest patterns.

First, for night lighting (SRQ2), the night-cool scene (C) was rated worse than the night-warm scenes (D/E) on all single items (Table 5.5; BH FDR ≤ .033). Consistent with this, Distress and Engagement were lower under warm lighting than under cool lighting at night. Median within-participant shifts were typically about one category on the 1–5 scales, indicating practically noticeable improvements when moving from cool to warm lighting within the night condition.

Second, day-versus-night contrasts (SRQ1) show that day scenes (A/B) are generally preferred over night scenes (C/E) on the same appraisals.

Third, tree effects (SRQ3) were smaller and less consistent in the rank tests. Where significant, they tended to accompany the warm-lighting gains rather than overturn the lighting pattern.

Friedman's omnibus screens showed significant, systematic within-participant ordering across the five scenes for every outcome (all $p < .001$). Concordance (Kendall's W) was highest for Safety and Use-Post ($W \approx 0.64$–0.69), moderate for Use and Attractiveness ($W \approx 0.38$–0.50), and smaller but reliable for Engagement and Distress ($W \approx 0.12$–0.21). Sensitivity analyses excluding $|z|>3$ outliers did not change this pattern. Taken together, the rank-based contrasts support a narrative of a night penalty at the station plaza, and warmer, more even night lighting yields robust gains in perceived Safety, Attractiveness, and intended Use, with concurrent reductions in Distress and Engagement relative to night-cool.

Table 5.4: Significant Wilcoxon results that passed BH-FDR

Outcome	Contrast	n	Median [95% CI]	p_FDR
Distress	C vs E	47	0.00 [-0.12, 0.00]	0.017
Engagement	C vs E	47	0.00 [-0.50, 0.00]	0.002
Attractive	B vs E	46	0.00 [-1.00, 0.00]	0.033
	C vs E	47	1.00 [1.00, 1.00]	<.001
Safe	B vs E	46	-1.00 [-1.00, 0.00]	<.001
	C vs E	47	1.00 [1.00, 1.00]	<.001
Use	B vs E	46	0.00 [-1.00, 0.00]	0.006
	C vs E	47	1.00 [0.00, 1.00]	<.001
Use post	B vs E	47	-1.00 [-1.00, 0.00]	<.001
	C vs E	47	1.00 [0.00, 1.00]	<.001

Table 5.5: Friedman's Omnibus screen, indicates the effect size (concordance) with Kendall's W. Was conducted for the main outcomes (left) and with sensitivity test (excluding SD |z|>3 outliers).

	Main				**Sensitivity**			
Outcome	**N**	**chi2**	**p**	**W**	**N**	**chi2**	**p**	**W**
Engagement	45	23.37	<.001	0.208	42	23.65	<.001	0.225
Distress	45	14	0.007	0.124	42	11.78	0.019	0.112
Attractive	46	57.31	<.001	0.498	43	51.87	<.001	0.483

Use	46	43.54	<.001	0.379	43	38.41	<.001	0.357
Safe	46	77.61	<.001	0.675	43	74.63	<.001	0.694
Use post	47	74.88	<.001	0.637	44	71.78	<.001	0.653

5.2.2 Mixed models (Linear Mixed Model, Cumulative Link Mixed Model)

Mixed models estimate scene effects while adjusting for covariates, random intercepts, and baseline differences. Results are consistent with the rank tests and show how strong each design factor's impact is. Continuous outcomes of the SSSQ were tested with LMM, as shown in Table 5.6. Nighttime (SRQ1) increased Distress (β ≈ +0.15, p = .003) and Engagement (β ≈ +0.56, p < .001) relative to day, while warm lighting (SRQ2) reduced both (Distress β ≈ -0.13, p = .007; Engagement β ≈ -0.39, p < .001). Trees (SRQ3) showed no effect on either subscale. Baseline subscale scores were positively associated with scene scores (as intended controls), whereas PSS-4 was not associated (all p > .13). These indicate an overall night penalty and a warm-lighting benefit on the psychological stress channel.

Table 5.6: LMM main effects

Outcome	Variable	β	CI low	CI high	p
Engagement	(Intercept)	0.665	-0.207	1.537	0.131
	Night	0.562	0.367	0.758	<.001
	Warm light	-0.392	-0.587	-0.197	<.001
	Trees	-0.01	-0.158	0.138	0.895
	Engagement baseline	0.590	0.330	0.85	<.001
	pss4	0.007	-0.082	0.097	0.865
Distress	(Intercept)	0.478	0.086	0.871	0.018
	Night	0.149	0.053	0.245	0.002
	Warm light	-0.133	-0.229	-0.037	0.006
	Trees	-0.019	-0.092	0.053	0.597
	Distress baseline	0.418	0.190	0.646	<.001
	Pss-4	0.040	-0.013	0.093	0.137

Single-item appraisals were tested with CLMM (Table 5.7), and shows that night sharply reduced odds of reporting a higher category (Safety OR ≈ 0.02; Attractiveness OR ≈ 0.06; Use OR ≈ 0.05; Use-Post OR ≈ 0.03; all *p* < .001). By contrast, warm lighting at night produced large gains (e.g., Attractiveness OR ≈ 9.76; Safety OR ≈ 8.68; Use OR ≈ 6.01; Use-Post OR ≈ 3.92; all *p* ≤ .001). Trees showed small, mixed main effects: a negative association for Use (OR ≈ 0.49, *p* = .025) and a borderline positive association for Use-Post (OR ≈ 1.81, *p* = .053); no clear effects on Safety/Attractiveness. PSS-4 was not associated with the ordinal appraisals (all $p \geq .18$). Proportional-odds checks via LR tests were non-significant for Night/Warm light/Trees, supporting a single OR per predictor.

Table 5.7: CLMM main effects

Outcome	Variable	OR	OR_low	OR_high	p
Safe	Night	0.0186	0.0074	0.0469	<.001
	Warm light	8.6774	3.938	19.12	<.001
	Trees	0.6455	0.3654	1.1402	0.1315
	pss4	1.0538	0.7845	1.4156	0.7276
Attractive	Night	0.0567	0.0248	0.1295	<.001
	Warm light	9.7565	4.4258	21.508	<.001
	Trees	1.1093	0.6367	1.9328	0.7142
	pss4	1.1985	0.9182	1.5642	0.1828
Use	Night	0.0465	0.0185	0.1165	<.001
	Warm light	6.0056	2.6278	13.725	<.001
	Trees	0.4899	0.2626	0.9142	0.025
	pss4	1.3258	0.8476	2.0739	0.2166
Use post	Night	0.0275	0.0109	0.0693	<.001
	Warm light	3.9185	1.807	8.4973	<.001
	Trees	1.8104	0.9922	3.3035	0.0531
	pss4	1.0406	0.7273	1.4887	0.8277

5.2.3 Moderation by gender and age

Moderation analyses indicate gender differences in the appraisal channel, while the moderation of the SSSQ outcomes was not observed (Tables 5.8 and 5.9). The night penalty for appraisals is clear, and the warm-lighting benefit is evident across several appraisals for both genders. Both women and men move in the same direction, but women show larger shifts, especially for safety and attractiveness.

Age interactions were small and imprecise. Multipliers per +10 years were near 1 with wide confidence intervals (for example, intended use under warm light: OR× ≈ 1.55 [0.66, 3.66]; Use under trees: OR× ≈ 1.10 [0.48, 2.50]; Use-Post under trees: OR× ≈ 1.74 [0.73, 4.17]). These patterns suggest that gender moderates some appraisal effects, while age does not show systematic moderation in this sample.

Table 5.8: Gender simple effects (Odds Ratio [95% CI]). ORs reflect the odds of a female participant giving a higher score in that condition compared to the reference condition/level, and similarly for a male participant.

Outcome	Context	Gender: Female OR [95% CI]	Gender: Male OR [95% CI]
Attractive	Night	0.04 [0.02, 0.12]	0.06 [0.02, 0.18]
	Warm light	5.62 [2.24, 14.05]	19.40 [6.16, 61.15]
	Trees	0.82 [0.39, 1.74]	1.30 [0.61, 2.78]
Safe	Night	0.01 [0.00, 0.03]	0.03 [0.01, 0.08]
	Warm light	13.40 [3.69, 48.69]	9.41 [3.88, 22.82]
Use-Post	Night	0.01 [0.00, 0.05]	0.03 [0.01, 0.21]
	Warm light	1.90 [0.61, 5.93]	9.92 [2.98, 32.97]

Trees	2.42 [0.98, 6.00]	1.50 [0.65, 3.46]

Table 5.9: Age moderation. OR Multipliers indicate the change in the odds ratio for every 10-year increase in participant age. Multipliers below 1 indicate a negative relationship; above 1, a positive relationship.

Outcome	Context	OR Multiplier per +10 years [95% CI]
Use	Night	1.40 [0.60, 3.31]
	Warm light	1.55 [0.66, 3.66]
	Trees	1.10 [0.48, 2.50]
Use-Post	Warm light	1.44 [0.68, 3.05]
	Trees	1.74 [0.73, 4.17]

5.3 Physiological outcomes

The physiological outcomes were noisier and challenging to interpret. The experiment is designed to allow participants to walk, while wrist-worn sensors are already less stable than questionnaires. After quality control, EDA contributed 173 windows, a scenario experienced by a participant, (70.3%) and HR contributed 188 windows (76.4%). This attrition is typical for short, movement-allowed VR windows and is addressed by the modeling strategy.

5.3.1 Planned contrasts: Wilcoxon

Physiological arousal (SCL, HR) did not differentiate scenes reliably under the short exposure windows of each scene, although SCL showed small and non-significant improvements under warm lighting. These results complement the clear psychological shifts, suggesting that appraisal changes occurred without systematic increases in tonic arousal (Table 5.10, Figure 5.2: Violin plots of planned contrasts with significance, per SRQ (left- SRQ1; right – SRQ3). Figure 5.2 – the violin plot). All planned pairs were tested (BH-FDR correction). Most contrasts are null after FDR. The night lighting contrast for SCL (C-D) shows a small negative median Δ (≈−0.055 µS, CI touches 0; FDR p = 0.275). The D-E contrast trends negative (≈−1.14 bpm; FDR p=0.672). No pair passed FDR.

Table 5.10: Wilcoxon's planned contrasts outcomes per SCL and HR.

DV	Contrast	SRQ	n	Median Δ	95% CI bootstrap	W	p	p BH-FDR
Skin Conductance Level (µS)	A–C	SRQ1 Day/Night	25	0.053	[−0.030, 0.410]	128.0	0.3666	0.4399
	B–E	SRQ1 Day/Night	24	0.007	[−0.051, 0.293]	114.0	0.3165	0.4399
	C–D	SRQ2 Lighting	28	−0.055	[−0.251, −0.003]	137.0	0.1375	0.2750
	C–E	SRQ2 Lighting	25	−0.117	[−0.469, 0.056]	92.0	0.0588	0.2750
	A–B	SRQ3 Trees	24	0.018	[−0.051, 0.064]	135.0	0.6840	0.6840
	D–E	SRQ3 Trees	28	−0.050	[−0.327, 0.028]	136.0	0.1315	0.2750
Heart Rate (bpm)	A–C	SRQ1 Day/Night	29	0.121	[−1.932, 2.987]	207.0	0.8314	0.8314
	B–E	SRQ1 Day/Night	30	1.528	[−1.582, 2.407]	188.0	0.3707	0.6777
	C–D	SRQ2 Lighting	30	2.038	[−0.782, 3.927]	153.0	0.1048	0.6290
	C–E	SRQ2 Lighting	29	0.573	[−1.661, 2.116]	190.0	0.5647	0.6777
	A–B	SRQ3 Trees	31	−0.319	[−2.871, 1.373]	217.0	0.5551	0.6777
	D–E	SRQ3 Trees	31	−1.142	[−3.034, −0.166]	185.0	0.2241	0.6723

Accordingly, Friedman tests did not detect reliable, systematic within-participant ordering across scenes for SCL or HR (SCL: $\chi^2(4)=8.08$, $p=.089$, $W=.065$, $n=31$; HR: $\chi^2(4)=8.34$, $p=.080$, $W=.074$, $n=28$). Concordance (effect) was small, indicating a weak and inconsistent ranking of scenes.

Table 5.11: Friedman's omnibus screen for physiological outcomes

Outcome	Participants	k (levels)	χ²(df=k−1)	p-value	Kendall's W effect size
SCL (µS, mean per window)	31	5	8.08 (df=4)	0.089	0.065
HR (bpm, mean per window)	28	5	8.34 (df=4)	0.080	0.074

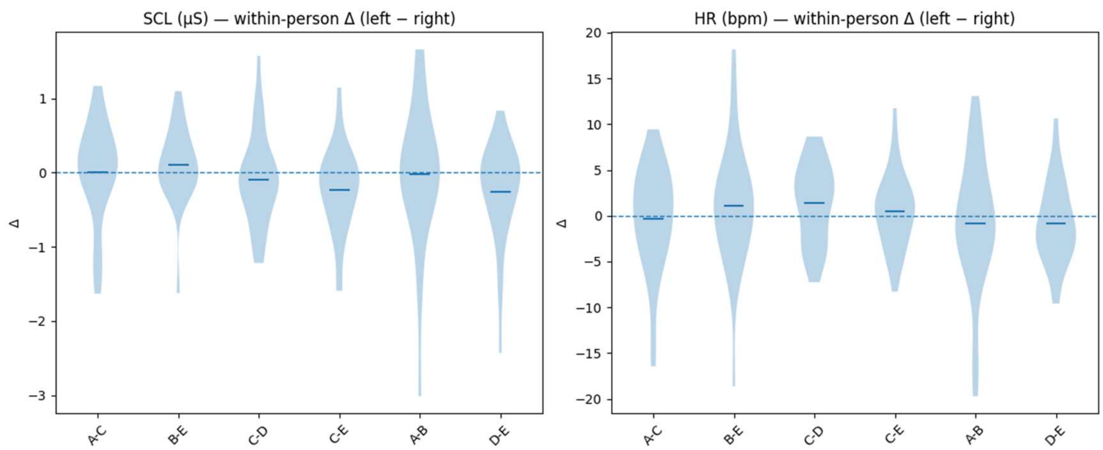

Figure 5.2: Violin plots of planned contrasts with significance, per SRQ (left- SRQ1; right – SRQ3).

5.3.2 Mixed models: Linear Mixed Model

The LMM demonstrates that scene attributes did not yield reliable physiological effects after accounting for repeated measures and covariates (Table 5.12). For HR, coefficients for Night (β ≈ −0.06, p = .956), Warm light (β ≈ −1.51, p = .125), and Trees (β ≈ +0.81, p = .272) were non-significant. For SCL, Night (β ≈ +0.03, p = .838), Warm light (β ≈ +0.02, p = .882), and Trees (β ≈ +0.03, p = .769) were likewise non-significant.

By contrast, participant-centered skin temperature was a robust positive covariate for SCL (β ≈ +0.54 µS/°C, p ≈ .001), validating its inclusion and indicating that small thermal shifts can masquerade as tonic arousal changes. The LMMs suggest that the apparent psychological shifts across scenes did not appear in consistent changes of physiology.

Table 5.12: Linear mixed model for physiological outcomes.

DV	Term	Estimate	SE	z	p
Heart Rate	Intercept	70.42	1.2444	56.591	0
	Night	-0.055	0.9933	-0.056	0.9557
	Warm light	-1.513	0.9863	-1.534	0.1251
	Trees	0.8097	0.7367	1.0991	0.2717

Skin Conductance Level	Intercept	2.0722	0.4172	4.9672	<0.001
	Night	0.0287	0.1404	0.2044	0.8381
	Warm light	0.0204	0.1375	0.1482	0.8822
	Trees	0.0314	0.1067	0.2941	0.7687
	Temperature	0.5401	0.1619	3.337	0.0008

5.4 Summary

Across the five scenes, psychological outcomes showed the clearest differences: at night, people rated the plaza as less safe, less attractive, and less worth using, while warm, more even lighting at night largely restored these ratings and reduced stress appraisals. Physiological indices (SCL, HR) were mainly unchanged. Table 5.13 summarizes the findings by SRQ. For SRQ1, night clearly worsens psychological outcomes, with physiology remaining stable. For SRQ2, enhanced warm night lighting improves appraisals and reduces perceived stress without increasing arousal. For SRQ3, trees do not harm perceived safety, and their effects are minor compared with lighting and not visible in physiology. For SRQ4, women benefit more from warm night lighting than men, while age effects are weak.

Table 5.13: Summarize map of the results. ↑ signifies a positive shift (↑↑/↑↑↑ for moderate/large shifts); ↓ is a negative shift; under brackets () is small or uncertain. – means no reliable effect was found.

SRQ/Variable	Single items				Continuous (SSSQ)		Sensors	
	Safety	**Attractiveness**	**Use**	**Use-Post**	**Distress**	**Engagement**	**SCL**	**HR**
SRQ1 Day/Night	↑↑	↑↑↑	↑↑	↑↑	↓	↓	-	-
SRQ2 Lighting	↑↑↑	↑↑↑	↑↑	↑	↓	↓	(↓)	-
SRQ3 Trees	-	-	↓	(↑)	-	-	-	-
SRQ4 Gender (Females benefit)	↑	↑	↑	↑	-	-	-	-

6 Discussion

In this chapter, key findings from the experiment are presented, then interpreted through the theoretical framework lenses of ethics and behavior, linked back to the research questions and hypotheses, compared with similar study findings, and translated into guidelines for public space design and future research, finished with a section of reflection over the thesis process.

6.1 Key findings

Across the five scenes, psychological outcomes were more robust and consistent, while physiological measures were mainly flat. Four patterns emerged:

1. **Night penalty.** Night scenarios produced lower ratings on safety, attractiveness, and intended use, together with higher distress, consistent with higher experienced stress at the plaza.
2. **Warm light benefits**. Lighting the plaza with the better-distributed warm light (D, E) reversed much of the night penalty, indicating a more legible plaza.
3. **Minor tree effect.** Trees did not harm perceived safety, and their influence on use was modest compared to lighting, but they did slightly improve perceptions of daytime.
4. **Gender differences.** Men and women moved in the same direction, but improvements under warm lighting were larger for women, suggesting uneven conversion at night can be reduced by lighting quality. Age did not show uniform moderation.

Tonic SCL and HR remained largely unchanged, which is expected under the experiment's conditions. Psychological questionnaires differentiated the scenes without a consistent tonic shift. In the questionnaires, single-item appraisals showed a stronger trend than distress and engagement (SSSQ).

6.2 Interpretation of the findings

The theoretical framework's lenses help connect observed shifts in scene-level questionnaires to plausible changes in immediate intention and further explain the effects of the findings and future trends that foster a sustainable and inclusive public space.

The ethical-spatial lens (see Section 2.1) asks whether the plaza supports real freedom to approach and use the metro without undue stress. A capability-supporting scene is one in which safety, attractiveness, and intended use increase, and distress does not increase (ideally decreases). It improves the wellbeing of pedestrians, and can suggest better future functionings, achievements.

The behavioral lenses (see Section 2.2) clarify how scene-level appraisals link design to immediate intention to use the metro, and which human-scale qualities matter most. Scene changes that improve safety and attractiveness and increase intended use among participants are consistent with the Attitude-Intention pathway at the station threshold, and can imply future behavioral change. Improved civility, by "respecting" the infrastructure provided to pedestrians, and improving the design features in the plaza, leads to a better quality of public space.

SRQ1: Night penalty. <u>Ethical:</u> After dark, the same physical plaza converts into less real freedom: Safety, Attractiveness, and Intended Use fall, and Distress rises. This is a capability contraction, especially for women, who experienced a larger night penalty. <u>Behavioral:</u> signals a loss of protection and civility, and a downward shift in attitude and intention. Especially at winter, when the night is longer, some commute both ways at dark. This affects stronger the decision to use or avoid the metro.

SRQ2: Warm lighting. Ethical: The plaza provides greater freedom at the same site and time without increasing physiological load by improving lighting quality. It acts as a capability-enlarging measure and offers greater accessibility for pedestrians, improved wellbeing, and future functionings. Behavioral: Restored protection improves comfort, and a higher intention to visit the plaza could, over time, support higher ridership. The metro plaza is promoted as a civil, safe space, with improved qualities.

SRQ3: Trees. Ethical: Trees offer secondary capability support. They add small improvements in some appraisals and do not create a physiological penalty. It may act as capability-enlarging, but at some scenarios at night, its effect has to be tested to prevent unwanted activities. Behavioral: Contributes mainly to comfort and enjoyment, especially at daytime, with no clear safety harm in this experiment.

SRQ4: Gender differences. Ethical: Larger night penalties and larger gains from warm lighting for women are a classic case of uneven conversion: the same resource yields less real freedom for one group. Better night lighting narrows this gap. Therefore, it is capability-expanding, improves wellbeing, and supports equality and inclusivity for groups who currently feel less protected. Behavioral: Improved lighting can reduce uneven conversion and support a more civil, human-scale design, which can by time change behavior and increase ridership of women, especially by night.

6.2.1 Related work

Other immersive VR experiments have shown similar within-participant shifts in pedestrian appraisals. Argota Sánchez-Vaquerizo et al. used a street-scenario experiment to compare future street layouts and collected self-reported questionnaires and physiological responses over stress, using SSSQ-10, similar to this study. Their work supports the logic used here: hold the place constant, manipulate micro-design attributes, and read most effects through subjective judgments. They found that subtle design changes can shift stress-related appraisals in public space street designs. Son et al. found that qualitative aspects of lighting strongly shape perceived safety after dark, and better night lighting reduces fear. Unlike this thesis, they emphasize the role of interior building lights along residential streets, which were not manipulated in this study. The relevant housing next to Bullewijk are distant away, which helps in isolating the lighting conditions of the plaza.

Lighting-focused reviews in VR further reinforce this approach. Scorpio et al. synthesize evidence on how uniformity and correlated color temperature shape reassurance, legibility, and comfort in outdoor settings. The night penalty observed in this thesis is consistent with VR and field studies showing that after-dark environments with low uniformity and obstructed sightlines elevate fear and degrade perceived quality, especially for women. Similarly, White et al. developed an immersive virtual e-participation method in which citizens evaluated alternative streetscape designs on perceived safety and place quality, showing that changes in elements such as lighting and greenery systematically shift reassurance and attractiveness. Bar-Ad et al. evaluated the addition of greenery in an urban VR setting and found that trees and planting improve perceived environmental quality and comfort.

Interestingly, Hedblom et al. used a 2D multisensory VR of visuals, sounds, and smells of urban green space. They found that park and forest environments produced lower SCL than dense urban environment. These findings do not contradict this thesis, but use larger sample and slightly different methodology, especially with the introduction of odors (Argota Sanchez-Vaquerizo et al., 2023; Bar-Ad et al., 2025; Ghanbari et al., 2024; Hedblom et al., 2019; Scorpio et al., 2020; Son et al., 2024; White et al., 2023).

6.3 Findings in relation to the research question and hypotheses

This section reviews the main research question, the four sub-research questions, and the associated hypotheses, and relates them to the empirical results from Chapter 5. The focus is on the psychological outcomes, where the clearest patterns emerged, with physiological outcomes used as context. The key findings are interpreted through theoretical lenses, and the hypotheses are verified.

SRQ1 and H1. Asked how stress differs between day and night variants of the plaza. H1 stated that participants would experience higher stress at night than during the day. The results support this expectation in the psychological channel. Compared with day scenes, the cool-light night scene, the baseline as-is, was rated as less safe, less attractive, and less likely to be used, while distress was higher and engagement was modestly elevated. In the physiological channel, SCL and HR did not show systematic day-night differences. Taken together, the findings indicate a robust night penalty in perceived stress and safety, while physiological arousal remained essentially unchanged at the time scale of the experiment

SRQ2 and H2. Examined how changes in nighttime lighting attributes affect stress, with and without trees. H2 proposed that warmer, better-distributed lighting would reduce stress compared to cooler lighting. The warm lighting scenes produced consistent and sizeable gains in psychological outcomes relative to the existing white lighting at night. Safety, attractiveness, and intended use increased, while distress decreased. Engagement did not rise in a way that suggested agitation, but instead matched a more focused, comfortable state. Physiological measures again showed no reliable differences across lighting variants. Overall, H2 was supported in the psychological channel, with lighting quality emerging as the main lever for reducing the night penalty.

SRQ3 and H3. Asked how the presence of trees changes stress across day and night. H3 anticipated that trees would tend to reduce stress, with possible dependence on time of day. The effects of trees were smaller and more mixed than those of lighting. In some appraisals, trees slightly improved attractiveness or use intentions, and they did not reduce perceived safety when combined with adequate lighting and open sightlines. At the same time, changes were modest, sometimes negligible, and absent in physiology. H3 therefore received only partial support: trees did not harm perceived safety and sometimes added comfort, but the dominant lever for reducing experienced stress remained lighting rather than vegetation.

SRQ4 and H4. Investigated whether the impact of lighting and trees on stress differed by gender and age. H4 expected differences in stress responses by gender and age, especially at night. The analyses showed that women experienced a larger night penalty than men and gained more from warm night lighting across several appraisals, suggesting that improved night lighting can narrow gender gaps in perceived safety and willingness to use the plaza. Age-related interactions were weaker and less consistent, and no clear pattern emerged in the physiological outcomes. Thus, H4 found support for gender-based differences in psychological responses, while age differences were limited in the current sample and design.

The main research question asked how adjustments in station plaza design, specifically nighttime lighting and the presence of trees, affect pedestrians' experienced stress at transit stations. It is answered using the four SRQs answered above. The results indicate that design changes at the entrance realm can meaningfully alter how the plaza is experienced. Low-cost lighting adjustments, combined with carefully placed trees, can reduce experienced stress at Bullewijk and make night scenes feel closer to day in terms of safety, perceived quality and willingness to use. By doing so, the plazas support a gender-inclusive and safer access to sustainable mobility, which improves the wellbeing of pedestrians, providing an infrastructure that respects them, and promotes public space qualities that set the intention for a behavioral change in society.

6.4 Recommendations

This section translates the findings into guidance for policy and design, both for the specific site of Bullewijk plaza and for similar entrance realms of public transportation. These recommendations are relevant to both policy, design, and approach for urban planning and design, including site-specific recommendations, in line with the site analysis (Section 3.2).

Lighting the plaza. Warm lighting with high distribution and lower strength was found helpful in improving the quality of the plaza. Upgrading the lighting is a simple, easy, and low-cost design tweak that dramatically benefits the well-being qualities of the public space, and the use of the metro. During winters, when night is longer, this is especially important.

Enhanced trees. Trees effects were small to mixed, and not harmful in night. However, according to other research and frameworks, trees can act as protection and enhance engagement in public space. Placing trees in other positioning as shown in the experiment might lead to different perception.

Stakeholder value. The VR environment proved useful as a dialogue tool with pedestrians, users, residents and practitioners. Lighting options can be evaluated with users, and can help with interaction with local communities for better planning and engagement. The questionnaires were robust and showed a trend.

Site-specific recommendations: Bullewijk Plaza. The plaza is situated at a crucial crossroads between two very different neighborhoods. It has the potential to act as an inclusion tool rather than a segregation tool. Simple tools, as introduced in this study, can significantly improve the plaza's qualities. As studied, the conditions are the worse during the night. Elevating the lighting in the plaza, distributing it better and assuring warmer lights is crucial. Although trees did not result in a significant improvement, they would not cause much harm even in night scenarios. Therefore, even though planting tall trees takes time, mobile pots can be placed to create a better public space. In addition, using the frameworks introduced in this thesis and adopting a tactical urbanism or placemaking approach, the further design of this plaza is highly beneficial while addressing the main cause of stress identified in this study: lighting at night. Other design patterns are possible, and with little investment, because of the plaza's underdeveloped nature, this public space could improve drastically, which in turn would act as a capabilities-expander, contributing to inclusivity and drawing changes in patterns of accessibility, ridership, and engagement in public space. It would be interesting to see on-site experiments or monitoring: with enhanced lighting, how does local surveys of safety, attractiveness and intended use change.

6.5 Limitations

This experiment was conducted as a master's thesis scope (as introduced in the introduction section) with limited capacity to recruit participants, no compensation for participants, and no budget for developing the virtual environment. Therefore, the nature of this experiment was limited to focus on mostly students who volunteered.

Construct Validity. The psychological outcomes of appraisals serve as proximal indicators, not long-run behavioral outcomes. The physiological arousal indices are under movement. It can undermine sensitivity and create noisy data. The phasic SCR and HRV were excluded because of these reasons. This research mitigated these problems by framing the appraisals as proximal indicators, and the SCL was controlled with temperature.

Internal validity. The short, walking-allowed windows increased ecological realism but produced noisier physiological data. The VR model was kept simple, with limited control over exact lighting color in kelvins. Trees were generalized, and the distributions of trees and lighting in the enhanced scenes were selected by the author and could have been chosen differently. Several participants reported that the place felt familiar, and some noted that it seemed cleaner in VR than in reality. Planned contrasts reduced fatigue effects and between-person noise, and sensitivity analyses with outlier removal preserved the main conclusions. Lighting was calibrated within the VR environment to approximate real-world luminance, color temperature, and shadow patterns.

Social validity. The scenes had no other pedestrians (although sounds of far-away activity was heard). Participants might infer which scenes are "supposed" to feel safer at night. These were mitigated with neutral instructions, a within-subject design, and the lack of other pedestrians was acknowledged within the scope of the research.

Sample and demographics. The sample size of this experiment (n = 47) and the characteristics of the participants is a limitation for this study. Participants were mostly students, educated and internationally diverse sample. No participant was actually living in the vicinity of Bullewijk. This research based on volunteers, and could not compensate other participants.

Scope. The manipulations were limited: only Night, Lighting and Trees, while other plausible levers such as seating, kiosks, litter and maintenance, other pedestrians and pavement were not tested. Further, the experiment was conducted in only one plaza with a specific lighting baseline. This can limit the generalization to other morphologies, and focus on the tested levers only.

6.6 Future research

Improving the experimental setup. To strengthen construct validity, future work can test alternative appraisals or questionnaires, add short stationary segments or longer exposure windows, and include devices better suited for stress detection under movement. This could make HRV and SCR usable alongside SCL and HR. In addition, to improve internal validity, varied exposure intervals and break lengths can be used to obtain cleaner physiological time series for each scenario, and refine calibration of lighting color and intensity. Furthermore, to address social validity, pedestrian flows with different characteristics can be added to examine how the presence and behavior of others change stress in plazas.

Extending the sample and scope. Future research should aim for larger, more diverse samples across age, gender, familiarity with the station, and socio-economic background. Including local residents and frequent users of Bullewijk, and possibly compensating participants, could increase both statistical power and social relevance. Moreover, expanding the scope to additional station plazas and to other design levers such as seating, kiosks, maintenance cues, and temporary installations would help clarify which combinations of interventions are most effective at reducing stress and elevating perceived safety. Comparative studies across multiple sites could test how robust the present findings are beyond Bullewijk.

Field validation and co-creation. As a site-specific next step, field validation and co-creation sessions can complement the VR experiment. On-site surveys at Bullewijk can establish a baseline and later track the impact of implemented interventions, for example, using the same single-item appraisals on safety, attractiveness, and intended use that were used in the VR study. These can be combined with simple visualizations or short VR clips shown to pedestrians at the station to probe reactions to proposed lighting and tree changes. Night site walks with women and older pedestrians, followed by structured feedback on different design options, can be especially valuable.

6.7 Reflection

This project required substantial effort to build the site-specific VR replica of the metro station, tune lighting assets, and recruit a diverse sample. It turned out to be a bigger journey than I expected, but it resulted in an environment that closely resembled the location, drawing positive responses from participants. In addition, the statistical analysis required much dedication, especially coming from a background of architecture, and turned out to be rewarding with robust results.

The clearest signals emerged in the psychological channel, while physiological metrics mainly remained flat. In hindsight, I underestimated the noise introduced by short, movement-allowed windows and wrist-worn sensing. HRV and SCR were especially sensitive to motion and failed QC more often than expected, which took plenty of time to discover and let go. I intentionally chose free walking movement for participants rather than a stationary or teleportation design for the VR experiment to improve realism and participant comfort. This introduced novelty and reliability for VR research, but it reduced detection power for physiological data. If designing the study again, I would plan the experiment differently to ensure fair physiological outcomes and consider using other devices to measure physiology. Despite these constraints, the within-participant design, planned contrasts, and linear mixed models produced a coherent pattern that is actionable for entrance-realm lighting decisions.

This research was my attempt to use an empirical, quantitative approach applied to a site-specific public space to understand how feasible changes can improve inclusivity, equality, accessibility, and opportunities. To promote a sustainable, safe, and low-carbon future, it is necessary to keep seeking high-quality entrance realms and improve the shared space.

7 Conclusions

This thesis examined whether design changes to lighting and trees at a metro station plaza can affect stress and the willingness to use the station. A virtual reality environment of Amsterdam's Bullewijk station plaza enabled within-participant exposure to five scene variants that crossed day versus night, with lighting color and distribution, and the presence of trees. Data was collected from forty-seven participants. Stress was assessed in two channels: psychological stress measurements (stress-state indices and appraisals) and physiological arousal indices (tonic Skin Conductance Level and Heart Rate). The study also tested for gender and age moderation.

To answer the research question, *'How can adjustments in station-plaza design, specifically nighttime lighting and the presence of trees, influence travelers' stress at transit stations?'*, the findings show warm and uniform night lighting at the entrance realm substantially improved psychological appraisals of safety, attractiveness, and willingness to use, and moderately improved distress, while tonic arousal did not rise. Trees had minor, context-dependent effects that were consistent with perceived safety, yet lighting remained the primary driver of improvement at night. The pattern was stronger for women, which experienced night as more stressful than men. In short, low-cost lighting upgrades at station plazas like Bullewijk can reduce perceived stress at night and the intention to use without a physiological cost, while careful tree placement can add comfort.

Across five scenes, psychological outcomes showed a robust pattern, while physiology remained essentially flat. First, a consistent night penalty emerged: relative to day, night scenes were rated less safe, less attractive, and less likely to be used, with higher distress and somewhat higher engagement. Second, warm night lighting produced sizeable improvements, reversing much of the night penalty without raising tonic arousal. Safety, attractiveness, and intended use increased strongly under warm lighting. Distress fell, and engagement did not increase in a way that would imply agitation. Third, the presence of trees showed small and mixed effects overall. Trees were compatible with perceived safety when sightlines were preserved, yet lighting remained the dominant lever at night. Fourth, moderation screens indicated gender-differential gains: women showed larger night penalties and larger improvements under warm lighting, suggesting that lighting quality can reduce uneven conversion of the same place into intention to use. Age effects were weak and imprecise over all scenarios and contrasts.

Read through a Capabilities approach, the entrance realm is part of the real freedom to move without fear. Warm and uniform lighting at night appears capability-expanding in this context because it raises Safety, Attractiveness, and Intended Use without a penalty in tonic arousal. The gender pattern points to equity implications: improving lighting quality can shrink capability gaps after dark. Gehl's human-scale criteria and Walker's civility reinforce the mechanism story. Finally, a Theory of Planned Behavior reading is consistent with the observed pattern that attitude-like appraisals relate to Intended Use at the scene level. The thesis does not claim long-run ridership effects, only a proximal intention link in context.

The research illustrates a site-specific VR protocol that holds the place constant while changing a small set of physical attributes that a municipality can realistically adjust. The protocol asked participants to actively walk, and had each scenario exposed for a short window of time. Therefore, psychological detection was clearer compared to the physiological outcomes.

The study used a single site with a specific under-viaduct morphology and a student-leaning sample. the project demonstrated a VR protocol that holds a real place constant while varying only a few adjustable physical features. Allowing natural movement and using short exposure windows supported psychological realism but reduced physiological sensitivity. Social context was simplified, with no other pedestrians or security presence in the scenes. These choices were deliberate to isolate lighting and trees.

This matters in a wider urban context. Amsterdam is pursuing a sustainable and low-car mobility system, in which the metro is a key component. For such a system to function equitably, the metro plaza must be a safe, pleasant, and inclusive space, one that both men and women feel comfortable using at all hours. Entrance realms are therefore not just access points but enablers of everyday mobility.

For station plazas with similar morphology, Night lighting quality is the highest priority. Warmer light at the threshold, even distribution of light, and allow no dark patches on the plaza. Trees are helpful when they preserve sightlines and improve the daytime experience. Gender-sensitive night audits during commissioning can help tune lighting to the users most affected by night penalties. The study shows that immersive VR can serve as a low-cost tool to compare lighting layouts before execution.

9 References

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Argota Sanchez-Vaquerizo, J., Hausladen, C. I., Mahajan, S., Matter, M., Siebenmann, M., A. B. Van Eggermond, M., & Helbing, D. (2023). A Virtual Reality Experiment to Study Citizen Perception of Future Street Scenarios. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4565113

Bar-Ad, R., Vigo, M., Caruso, G., Quboa, Q., & Pinto, N. (2025). Evaluating the impact of added greenery on perceived factors of an urban environment in virtual reality. *PLOS ONE*, *20*(2), e0316195. https://doi.org/10.1371/journal.pone.0316195

Biggs, A., Brough, P., & Drummond, S. (2017). Lazarus and Folkman's Psychological Stress and Coping Theory. In *The Handbook of Stress and Health* (pp. 349–364). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118993811.ch21

Brambilla, E., Petersen, E., Stendal, K., Sundling, V., MacIntyre, T. E., & Calogiuri, G. (2024). Effects of immersive virtual nature on nature connectedness: A systematic review and meta-analysis. *DIGITAL HEALTH*, *10*, 20552076241234639. https://doi.org/10.1177/20552076241234639

Bzdúšková, D., Marko, M., Hirjaková, Z., Kimijanová, J., Hlavačka, F., & Riečanský, I. (2022). The Effects of Virtual Height Exposure on Postural Control and Psychophysiological Stress Are Moderated by Individual Height Intolerance. *Frontiers in Human Neuroscience*, *15*. https://doi.org/10.3389/fnhum.2021.773091

Chen, S., Lin, S., Yao, Y., & Zhou, X. (2024). Urban Public Space Safety Perception and the Influence of the Built Environment from a Female Perspective: Combining Street View Data and Deep Learning. *Land*, *13*(12), 2108. https://doi.org/10.3390/land13122108

Du, J., Druta, O., van den Berg, P., & van Wesemael, P. J. V. (2021). How Do Socio-Demographic Characteristics Affect Users' Perception of Place Quality at Station Areas? Evidence from Amsterdam, The Netherlands. *Urban Science*, *5*(4), 80. https://doi.org/10.3390/urbansci5040080

Epel, E. S., Crosswell, A. D., Mayer, S. E., Prather, A. A., Slavich, G. M., Puterman, E., & Mendes, W. B. (2018). More than a feeling: A unified view of stress measurement for population science. *Frontiers in Neuroendocrinology*, *49*, 146–169. https://doi.org/10.1016/j.yfrne.2018.03.001

Feng, Y., Duives, D., Daamen, W., & Hoogendoorn, S. (2021). Data collection methods for studying pedestrian behaviour: A systematic review. *Building and Environment*, *187*, 107329. https://doi.org/10.1016/j.buildenv.2020.107329

Fotios, S., Unwin, J., & Farrall, S. (2015). Road lighting and pedestrian reassurance after dark: A review. *Lighting Research & Technology*, *47*(4), 449–469. https://doi.org/10.1177/1477153514524587

Gehl, J. (2013). *Cities for People*. Island Press.

Ghanbari, M., Dijst, M., McCall, R., & Perchoux, C. (2024). The use of Virtual Reality (VR) to assess the impact of geographical environments on walking and cycling: A systematic literature review. *International Journal of Health Geographics*, *23*(1), 15. https://doi.org/10.1186/s12942-024-00375-6

Hamilton-Baillie, B. (2008). Shared Space: Reconciling People, Places and Traffic. *Built Environment*, *34*(2), 161–181. https://doi.org/10.2148/benv.34.2.161

Hedblom, M., Gunnarsson, B., Iravani, B., Knez, I., Schaefer, M., Thorsson, P., & Lundström, J. N. (2019). Reduction of physiological stress by urban green space in a multisensory virtual experiment. *Scientific Reports*, *9*(1), 10113. https://doi.org/10.1038/s41598-019-46099-7

Helton, W. S. (2004). Validation of a Short Stress State Questionnaire. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *48*(11), 1238–1242. https://doi.org/10.1177/154193120404801107

Johansson, S., & Haandrikman, K. (2023). Gendered fear of crime in the urban context: A comparative multilevel study of women's and men's fear of crime. *Journal of Urban Affairs*, *45*(7), 1238–1264. https://doi.org/10.1080/07352166.2021.1923372

Kardan, O., Gozdyra, P., Misic, B., Moola, F., & Palmer, L. J. (2016). Neighborhood Greenspace and Health in a Large Urban Center. In *Urban Forests*. Apple Academic Press.

Knaap, G., & Talen, E. (2005). New Urbanism and Smart Growth: A Few Words from the Academy. *International Regional Science Review*, *28*(2), 107–118. https://doi.org/10.1177/0160017604273621

Lazarus, R. S., & Folkman, S. (1984). *Stress, Appraisal, and Coping*. Springer Publishing Company.

Li, L., & Lange, K. W. (2023). Assessing the Relationship between Urban Blue-Green Infrastructure and Stress Resilience in Real Settings: A Systematic Review. *Sustainability*, *15*(12), 9240. https://doi.org/10.3390/su15129240

Llaguno-Munitxa, M., Edwards, M., Grade, S., Vander Meulen, M., Letesson, C., Sierra, E. A., Altomonte, S., Lacroix, E., Bogosian, B., Kris, M., & Macagno, E. (2022). Quantifying stress level reduction induced by urban greenery perception. *IOP Conference Series: Earth and Environmental Science*, *1122*(1), 012021. https://doi.org/10.1088/1755-1315/1122/1/012021

Ma, S., Zhang, W., Noland, R. B., Andrews, C. J., Younes, H., & von Hagen, L. A. (2025). *Assessing Pedestrian Stress with Biometric Sensing and Survey Responses* (SSRN Scholarly Paper No. 5158338). Social Science Research Network. https://doi.org/10.2139/ssrn.5158338

Nussbaum, M. (2003). Capabilities as Fundamental Entitlements: Sen and Social Justice. *Feminist Economics*, *9*(2–3), 33–59. https://doi.org/10.1080/1354570022000077926

Nussbaum, M. C. (2005). Women's Bodies: Violence, Security, Capabilities. *Journal of Human Development*, *6*(2), 167–183. https://doi.org/10.1080/14649880500120509

Park, W., Son, D., Im, B., Her, J., Kim, Y.-J., & Kim, S.-N. (2025). Fear of crime revisited: Analyzing the effects of urban nighttime illuminance on neural and psychological responses using recorded virtual environments and electroencephalogram data. *Virtual Reality*, *29*(3), 145. https://doi.org/10.1007/s10055-025-01211-3

Rabaza, O., Peña-García, A., Pérez-Ocón, F., & Gómez-Lorente, D. (2013). A simple method for designing efficient public lighting, based on new parameter relationships. *Expert Systems with Applications*, *40*(18), 7305–7315. https://doi.org/10.1016/j.eswa.2013.07.037

Robeyns, I. (2005). The Capability Approach: A theoretical survey. *Journal of Human Development*, *6*(1), 93–117. https://doi.org/10.1080/146498805200034266

Robeyns, I., & Byskov, M. F. (2011). *The Capability Approach*. https://plato.stanford.edu/entries/capability-approach/?trk=public_post_comment-text

Rossetti, T., Lobel, H., Rocco, V., & Hurtubia, R. (2019). Explaining subjective perceptions of public spaces as a function of the built environment: A massive data approach. *Landscape and Urban Planning*, *181*, 169–178. https://doi.org/10.1016/j.landurbplan.2018.09.020

Savini, F., Boterman, W. R., van Gent, W. P. C., & Majoor, S. (2016). Amsterdam in the 21st century: Geography, housing, spatial development and politics. *Cities*, *52*, 103–113. https://doi.org/10.1016/j.cities.2015.11.017

Schönfeld, K. C. von, & Bertolini, L. (2017). Urban streets: Epitomes of planning challenges and opportunities at the interface of public space and mobility. *Cities*, *68*, 48–55. https://doi.org/10.1016/j.cities.2017.04.012

Scorpio, M., Laffi, R., Masullo, M., Ciampi, G., Rosato, A., Maffei, L., & Sibilio, S. (2020). Virtual Reality for Smart Urban Lighting Design: Review, Applications and Opportunities. *Energies*, *13*(15), 3809. https://doi.org/10.3390/en13153809

Sen, A. (1999). Commodities and Capabilities. *OUP Catalogue*. https://ideas.repec.org//b/oxp/obooks/9780195650389.html

Sen, A. (2008). The Idea of Justice1. *Journal of Human Development*, *9*(3), 331–342. https://doi.org/10.1080/14649880802236540

Son, D., Im, B., Her, J., Park, W., Kang, S.-J., & Kim, S.-N. (2024). Street lighting environment and fear of crime: A simulated virtual reality experiment. *Virtual Reality*, *29*(1), 8. https://doi.org/10.1007/s10055-024-01080-2

Stress. (n.d.). Retrieved November 18, 2025, from https://www.who.int/news-room/questions-and-answers/item/stress

Twohig-Bennett, C., & Jones, A. (2018). The health benefits of the great outdoors: A systematic review and meta-analysis of greenspace exposure and health outcomes. *Environmental Research*, *166*, 628–637. https://doi.org/10.1016/j.envres.2018.06.030

Tyrväinen, L., Ojala, A., Korpela, K., Lanki, T., Tsunetsugu, Y., & Kagawa, T. (2014). The influence of urban green environments on stress relief measures: A field experiment. *Journal of Environmental Psychology*, *38*, 1–9. https://doi.org/10.1016/j.jenvp.2013.12.005

Vecchio, G., & Martens, K. (2021). Accessibility and the Capabilities Approach: A review of the literature and proposal for conceptual advancements. *Transport Reviews*, *41*(6), 833–854. https://doi.org/10.1080/01441647.2021.1931551

Walker, J. (2024). *Human transit, revised edition: How clearer thinking about public transit can enrich our communities and our lives*. Island Press.

https://books.google.com/books?hl=en&lr=&id=KCbqEAAAQBAJ&oi=fnd&pg=PP1&dq=info:xd55CAUD7zsJ:scholar.google.com&ots=iTJszsdDnC&sig=dlUR7ByytBIy7mBcK5EaeYqk4ms

Wan, J., Wu, H., Collins, R., Deng, K., Zhu, W., Xiao, H., Tang, X., Tian, C., Zhang, C., & Zhang, L. (2024). Integrative analysis of health restoration in urban blue-green spaces: A multiscale approach to community park. *Journal of Cleaner Production*, *435*, 140178.

White, M., Langenheim, N., Yang, T., & Paay, J. (2023). Informing Streetscape Design with Citizen Perceptions of Safety and Place: An Immersive Virtual Environment E-Participation Method. *International Journal of Environmental Research and Public Health*, *20*(2), 1341. https://doi.org/10.3390/ijerph20021341

Yang, S., Dane, G., van den Berg, P., & Arentze, T. (2024). Influences of cognitive appraisal and individual characteristics on citizens' perception and emotion in urban environment: Model development and virtual reality experiment. *Journal of Environmental Psychology*, *96*, 102309. https://doi.org/10.1016/j.jenvp.2024.102309

Yu, C.-P., Lee, H.-Y., & Luo, X.-Y. (2018). The effect of virtual reality forest and urban environments on physiological and psychological responses. *Urban Forestry & Urban Greening*, *35*, 106–114. https://doi.org/10.1016/j.ufug.2018.08.013

Zhang, Y., Chen, G., He, Y., Jiang, X., & Xue, C. (2022). Social Interaction in Public Spaces and Well-Being among Elderly Women: Towards Age-Friendly Urban Environments. *International Journal of Environmental Research and Public Health*, *19*(2), 746. https://doi.org/10.3390/ijerph19020746

AlleCijfers.nl. (n.d.). *Duidelijke informatie in cijfers en grafieken*. AlleCijfers.nl. Retrieved October 20, 2025, from https://allecijfers.nl/

World Health Organization. (2023, February 21). *Stress*. World Health Organization. https://www.who.int/news-room/questions-and-answers/item/stress World Health Organization

10 Appendices

Appendix A: Data Cleaning Protocol

Data Cleaning Protocol: Physiological Data

All physiological streams (EDA, HR, temperature, ACC) were exported with Empatica Manager (CareLab), time-stamped in UTC, aligned to scenario start times, sampling rate, and per-second coverage were verified, sliced to participant-by-scene windows, summarized per participant, and resampled to 1 Hz. Cleaning was performed at native sampling rates. The outcomes were mean HR, mean SCL, and the temp-centered part, all presented in the master model table.

EDA. EDA was sampled at 4 Hz. Tonic-phasic decomposition was attempted when available using cvxEDA via NeuroKit (convex optimization). Phasic detections were frequently unreliable due to frequent movement. SCR counts and phasic AUC were explored but not retained, due to a low amount of detections. The tonic SCL mean was retained as the EDA outcome, labeled mean_SCL. EDA was cleaned with robust spike and flatline masks, short-gap interpolation, no ACC masking.

EDA cleaning. Three cleaning rules for EDA were performed. First, spike mask: EDA spikes were masked using a robust first-difference rule, where |*delta* EDA| exceeded approximately 3 to 4 times the median absolute deviation (MAD). Second, flatline mask: Near-flat segments were masked using a small amplitude floor of about 0.001 to 0.01 µS, while standard-deviation flatline checks were disabled. Third, interpolation: short gaps up to 3 seconds were interpolated. Longer gaps were left missing.

BVP (and HR). BVP was sampled at 64 Hz via photoplethysmography. Inter-beat intervals (IBI, ms) were available from device peak detection. HR was summarized as the scene-mean bpm and retained for analysis. HRV metrics (RMSSD, SDNN, lnRMSSD) were not analyzed due to high correction burden.

HR cleaning. Abnormal HR beats were flagged and treated according to the following rules: (i) out of physiological bounds (300-2000 ms) were removed; (ii) missed/double detections (two adjacent IBIs ≈ 2× local median) were fixed; (iii) rolling-median outliers (> 25% from the previous ~11 valid IBIs) were removed; (iv) ectopic beats excluded; (v) short gaps of RR bridged by interpolation, with %RR-corrected.

Temperature. Skin temperature was centered within-participant (scene temperature minus that participant's mean across scenes) to capture within-person thermal drift, for the SCL part. A fixed-effects regression of SCL on the participant-centered temperature showed a positive within-participant association ($\beta \approx 0.56$ µS/°C, cluster-robust SE, $p \approx 0.03$), corroborated by a mixed model with a participant random intercept ($\beta \approx 0.56$ µS/°C, $p < 0.001$).

Accelerometer. ACC was sampled at 32 Hz, transformed to vector magnitude, aligned, and averaged per scene to mean_ACC. ACC served descriptively to confirm that participants were moving during windows. Movement was near 100 across scenes, so ACC was not used as an artifact mask for EDA.

Missing data and handling. For EDA and temperature, gaps of 2s or less were linearly interpolated; longer gaps were left missing. For HR, isolated missing beats of up to 3 consecutive beats were replaced by spline interpolation on beat times; longer voids were treated as missing segments. Scene means were computed only if the scene passed QC. Otherwise, that dependent variable was marked invalid for that scene.

Baseline decision. A familiarization window was initially considered as a baseline to form per-scenario deltas (delta HR, delta SCL). This window frequently failed QC and showed unstable regulation, so deltas were dropped. Per-scenario means were adopted as the primary outcomes.

Sensitivity and robustness analyses. Several tests were conducted to ensure the quality of the data. First, varying the %RR corrected threshold between 10 and 30 did not alter the inferences for HR. HRV indices remained unstable and were not analyzed. Second, a comparison was made between QC-strict and QC-lenient (coverage thresholds). The QC-lenient results were in the same direction; therefore, the QC-strict results were reported. Third, means vs delta showed consistent results, but reported means (see above).

Data Cleaning Protocol: Psychological Data

First, all questionnaires were imported in R, then scored according to their keys. For the SSSQ-10 Engagement and Distress subscales, means were computed as the mean of their respective items. Missingness was near zero, so there was no need to execute the Little MCAR test. Cook's D was screened but not used, as within-scene z was the outlier detection. Second, data was reshaped to a long dataset with one row per participant per scene, carrying the derived Night/Yellow/Trees factors and the baseline. Third, covariates for moderation were labeled for SRQ4: center PSS-4, Age, and Gender. Fourth, analysis-ready wide and long tables were saved, followed by data checks, analysis tests, and sensitivity.

Appendix B: Informed Consent Form

Informed Consent Form (ICF) | Transit Stations' Plaza VR Experiment

You are invited to participate in a research study titled Transit Stations' Plaza. This study is being conducted by Yuval Tshuva from the AMS Institute's MADE program under the supervision of TU Delft and Wageningen University & Research.

The purpose of this research study is to explore public spaces outside metro stations in Amsterdam, which is expected to take approximately 45 minutes to complete. The data will be used for a master's thesis, which will be published in the AMS and TU Delft repository.

We will be asking you to wear the VR headset (Oculus Quest) and follow the researcher's instructions, followed by a questionnaire. You will participate in a Virtual Reality (VR) session lasting approximately 90 seconds each. During the session, you will experience different environments in a public space. After each scenario, you will complete a short questionnaire about your experience, focusing on stress and comfort levels. Biometric data, including heart pulse and skin conductance, will also be collected during the experiment.

While there may be no direct personal compensation for participating in this experiment, your participation will contribute to important research on urban design and public space.

As with any online activity, the risk of a breach is always possible. To the best of our ability, your responses to this study will be kept confidential. We will minimize any risks by storing all data confidentially, anonymously, and separately from the contact information, which will never be published. Only the researcher and the academic supervisor will have access to the data. Supervisors will not have access to sensitive personal data, including names and email addresses. The supervision will be involved with the data collected from the experiment's questionnaires and sensors.

Participation in this study is entirely voluntary, and you may withdraw at any time. You are free to omit any questions.

Responsible Investigator – Yuval Tshuva yuval.tshuva@wur.nl

Academic Supervisor – Dr. Yan Feng, Faculty of Civil Engineering y.feng@tudelf.nl

Consent

By signing below, you acknowledge that you have read, understood, and agreed to the information above. You agree to participate in the study voluntarily. You are at least 18 years old. You understand you can withdraw at any time.

Participant's Name: ____ #________

Participant's Signature: ______________________________ Date: ______________

Researcher's Name: Yuval Tshuva

Researcher's Signature: ______________________________ Date: ______________

Appendix C: Participant's briefing

<u>VR Experiment Instructions</u> – Public Space Outside A Metro Station in Amsterdam

Dear Participant,

Thank you for participating in this experiment, which examines the perception and stress levels experienced in public space settings. This research is focused on the plaza outside a metro station in Amsterdam.

This experiment uses an immersive Virtual Reality experience, which requires you to wear a headset and follow five scenarios. Each scenario lasts up to 90 seconds. After each scenario, you will be asked to complete a brief questionnaire. After the fifth scenario, you will be asked to fill in another short and final questionnaire. Please try to answer the questionnaires with care and attention, as they were kept short and not excessive. Although some may seem repetitive, they are carefully thought out.

Additionally, a wristband worn by the researcher will record your heartbeat and skin conductance, which will also be analyzed.

You may walk in a very selected and small area, which is marked both in the virtual world and in the lab.

The VR headset may cause some dizziness at one point, which is normal. However, this experiment is entirely voluntary, and you may stop at any point.

By participating in this research, you not only help me complete my master's thesis but also contribute to the improvement of public space. Please feel free to ask any questions regarding the study at the end of the session.

Please enjoy some coffee, cookies, and snacks at the end of the experiment.

Thank you for your time and patience, and thank you for supporting my research.

Appendix D: Participant's questionnaires: Demographics, Travel Habits

Demographics

Dear participant, thank you for attending this experiment. The following information is collected for the purpose of this study, and will not be identified or shared. This information helps us understand better the background of your choices and perceptions, and is highly important for understanding the results. Thank you for your understanding.

Please answer this and the following questionnaire to your best understanding.

What is your Participant Identification Number? #

What sequence code were you given by the researcher? (5 letters)

What is your age?

I identify as

- ◯ Male
- ◯ Female
- ◯ Non-binary / third gender
- ◯ Prefer not to say

Where do you currently live? Please enter the four first numbers of your postal code

In which country were you raised?

How do you define the neighborhood YOU GREW UP IN?

- ◯ City center / downtown
- ◯ Compact walkable neighborhood (outside the city center)
- ◯ Low density suburb
- ◯ Rural area / village

How do you define your CURRENT neighborhood?

- ◯ City center / downtown
- ◯ Compact walkable neighborhood (outside the city center)
- ◯ Low density suburb
- ◯ Rural area / village

How often do you experience a Virtual Reality (VR) environment (e.g. VR gaming, training, entertainment)?

- ◯ Very often
- ◯ Often
- ◯ Sometimes
- ◯ Seldom
- ◯ Never

What is your study/profession background?

- ◯ MADE Student
- ◯ Built environment, Architecture, Urbanism, Mobility
- ◯ Other

Do you have a driving license?

- ◯ Yes
- ◯ No

Do you own a car?

- ◯ Yes
- ◯ No, but I have access to the car of a family member or a friend
- ◯ No, but I often use car sharing or rental cars
- ◯ No

How many kilometers do you drive per week on average?

- ◯ 0
- ◯ <50
- ◯ 50-200
- ◯ 200-400
- ◯ >400

Please select which of the following modes of transportation you use REGULARLY (on weekly basis)

- ☐ Walking (as a mode of transportation)
- ☐ Bike
- ☐ BTM: Bus, Tram, Metro
- ☐ Train
- ☐ Motorbike
- ☐ Car / Shared car

How often do you use public transportation?

- ◯ Very often
- ◯ Often
- ◯ Sometimes
- ◯ Seldom
- ◯ Never

Do you live near the Bullewijk metro station? (<2 Kilometers)

- ◯ Yes
- ◯ No

Are you familiar with the Bullewijk metro station and its surroundings?

- ◯ Visiting daily
- ◯ Visiting weekly
- ◯ I have been there a few times
- ◯ I have been there once
- ◯ Never been

How often do you use the metro service in Amsterdam?

- ◯ Very often
- ◯ Often
- ◯ Sometimes
- ◯ Seldom
- ◯ Never

Have you ever experienced a situation in the Netherlands' metro system where you felt severely unsafe, threatened, or were a victim of robbery or assault?

- ◯ Yes
- ◯ No
- ◯ Yes, but outside of The Netherlands

Please describe it briefly

Appendix E: SSSQ, Single Items, PSS-4

SSSQ - B

This was the first scenario of the VR environment

Please indicate how well each word describes how you feel *At The Moment*.

	Not at all	A little bit	Somewhat	Very much	Extremely
Dissatisfied	○	○	○	○	○
Alert	○	○	○	○	○
Depressed	○	○	○	○	○
Sad	○	○	○	○	○
Active	○	○	○	○	○
Impatient	○	○	○	○	○
Annoyed	○	○	○	○	○
Angry	○	○	○	○	○
Irritated	○	○	○	○	○
Grouchy	○	○	○	○	○

Please select how relevant these statements to your experience in the previous scenario

	Not at all	A little bit	Somewhat	Very much	Extremely
I found the previous scenario attractive as a pedestrian	○	○	○	○	○
I would be willing to use the metro given the previous scenario	○	○	○	○	○
I felt safe in the previous scenario	○	○	○	○	○

PSS-4

The questions in this scale ask you about your feelings and thoughts during THE PAST MONTH.
In each case, please indicate your response by selecting the option representing HOW OFTEN you felt or thought a certain way.

In the last month, how often have you felt that you were unable to control the important things in your life?

Never	Almost Never	Sometimes	Often	Very Often
○	○	○	○	○

In the last month, how often have you felt confident about your ability to handle your personal problems?

Never	Almost Never	Sometimes	Often	Very Often
○	○	○	○	○

In the last month, how often have you felt that things were going your way?

Never	Almost Never	Sometimes	Often	Very Often
○	○	○	○	○

In the last month, how often have you felt difficulties were piling up so high that you could not overcome them?

Never	Almost Never	Sometimes	Often	Very Often
○	○	○	○	○

Appendix F: Post-questionnaire

Given these scenarios, please mark how likely are you to use the metro

	1 (Never)	2	3	4	5 (Very likely)
A - Daytime, clear	○	○	○	○	○
B - Daytime, with more trees	○	○	○	○	○
C - Nighttime, clear (Whiter Light)	○	○	○	○	○
D - Nighttime, more poles (yellow light)	○	○	○	○	○
E - Nighttime, more poles, more trees	○	○	○	○	○

Feedback:
Would you like to add a comment about the experiment? Please write it down here, thank you

Would you like to receive the results of this research? If you do, please fill in your email address, to which I will send a brief results section of my thesis. Thank you for participating.

Appendix G: Administered questionnaire: SSQ, iPQ

SSQ

For each symptom listed below, please rate its intensity based on how you feel **right now**

	None	Slight	Moderate	Severe
General Discomfort	○	○	○	○
Fatigue	○	○	○	○
Headache	○	○	○	○
Eye Strain	○	○	○	○
Difficulty Focusing	○	○	○	○
Increased Salivation	○	○	○	○
Sweating	○	○	○	○
Nausea	○	○	○	○
Difficulty Concentrating	○	○	○	○
Fullness of Head	○	○	○	○
Blurred Vision	○	○	○	○
Dizziness (eyes open)	○	○	○	○
Dizziness (eyes closed)	○	○	○	○
Vertigo	○	○	○	○
Stomach Awareness	○	○	○	○
Burping	○	○	○	○

iPQ

Now you'll see some statements about experiences. Please indicate, whether or not each statement applies to your experience. If a question is not relevant to the virtual environment you used, just skip it. You can use the whole range of answers. There are no right or wrong answers, only your opinion counts.
You will notice that some questions are very similar to each other. This is necessary *for statistical reasons* . And please remember: Answer all these questions only referring to this *one* experience.

In the computer generated world I had a sense of "being there"

-3 (Not at all)	-2	-1	0	+1	+2	+3 (Very much)
○	○	○	○	○	○	○

How aware were you of the real world surrounding while navigating in the virtual world? (i.e. sounds, room temperature, other people, etc.)?

-3 (extremely aware)	-2	-1	0	+1	+2	+3 (not aware at all)
○	○	○	○	○	○	○

How real did the virtual world seem to you?

-3 (completely real)	-2	-1	0	+1	+2	+3 (not real at all)
○	○	○	○	○	○	○

I had a sense of acting in the virtual space, rather than operating something from outside.

-3 (fully disagree)	-2	-1	0	+1	+2	+3 (fully agree)
○	○	○	○	○	○	○

How much did your experience in the virtual environment seem consistent with your real world experience ?

-3 (not consistent)	-2	-1	0	+1	+2	+3 (very consistent)
○	○	○	○	○	○	○

I was completely captivated by the virtual world.

-3 (fully disagree)	-2	-1	0	+1	+2	+3 (fully agree)
○	○	○	○	○	○	○

I did not feel present in the virtual space.

-3 (did not feel)	-2	-1	0	+1	+2	+3 (felt present)
○	○	○	○	○	○	○

Somehow I felt that the virtual world surrounded me.

-3 (fully disagree)	-2	-1	0	+1	+2	+3 (fully agree)
○	○	○	○	○	○	○

I was not aware of my real environment.

-3 (fully disagree)	-2	-1	0	+1	+2	+3 (fully agree)
○	○	○	○	○	○	○

I felt present in the virtual space.

-3 (fully disagree)	-2	-1	0	+1	+2	+3 (fully agree)
○	○	○	○	○	○	○

I still paid attention to the real environment.

-3 (fully disagree)	-2	-1	0	+1	+2	+3 (fully agree)
○	○	○	○	○	○	○

The virtual world seemed more realistic than the real world.

-3 (fully disagree)	-2	-1	0	+1	+2	+3 (fully agree)
○	○	○	○	○	○	○

I felt like I was just perceiving pictures.

-3 (fully disagree)	-2	-1	0	+1	+2	+3 (fully agree)
○	○	○	○	○	○	○

Appendix H: Ethical approval

Date 10-Feb-2025
Correspondence hrec@tudelft.nl

TU Delft Delft University of Technology

Human Research Ethics
Committee TU Delft
(http://hrec.tudelft.nl)

Visiting address
Jaffalaan 5 (building 31)
2628 BX Delft

Postal address
P.O. Box 5015 2600 GA Delft
The Netherlands

Ethics Approval Application: Transit Station's Plaza: How does lighting and greenery impact travelers' perception
Applicant: Tshuva, Yuval

Dear Yuval Tshuva,

It is a pleasure to inform you that your application mentioned above has been approved.

Thanks very much for your submission to the HREC which has been approved.

In addition to any specific conditions or notes, the HREC provides the following standard advice to all applicants:
• In light of recent tax changes, we advise that you confirm any proposed remuneration of research subjects with your faculty contract manager before going ahead.
• Please make sure when you carry out your research that you confirm contemporary covid protocols with your faculty HSE advisor, and that ongoing covid risks and precautions are flagged in the informed consent - with particular attention to this where there are physically vulnerable (eg: elderly or with underlying conditions) participants involved.
• Our default advice is not to publish transcripts or transcript summaries, but to retain these privately for specific purposes/checking; and if they are to be made public then only if fully anonymised and the transcript/summary itself approved by participants for specific purpose.
• Where there are collaborating (including funding) partners, appropriate formal agreements including clarity on responsibilities, including data ownership, responsibilities and access, should be in place and that relevant aspects of such agreements (such as access to raw or other data) are clear in the Informed Consent.

Good luck with your research!

Sincerely,

Dr. Ir. U. Pesch
Chair HREC
Faculty of Technology, Policy and Management

Appendix I: Data Management Plan

0. Administrative questions

1. Name of data management support staff consulted during the preparation of this plan.

My faculty data steward, Xinyan Fan, reviewed this DMP on 10.12.2024.

2. Date of consultation with support staff.

2024-12-10

I. Data description and collection or re-use of existing data

3. Provide a general description of the type of data you will be working with, including any re-used data:

Type of data	File format(s)	How will data be collected (for re-used data: source and terms of use)?	Purpose of processing	Storage location	Who will have access to the data
Questionnaires SSSQ (Short state stress)	.csv	Tablet, in the VR experiment lab (AMS Institute, Amsterdam) During the experiment - questionnaire	Understand participant's subjective reactions to different interventions in the public space	OneDrive	The research team: the PI and three professors: dr. Yan Feng (TU Delft); dr. Jaime Soza Parra (UU); prof. Alexander Klippel (WUR)
Physical data of participants - pulse rate, skin conductance, eye tracking	.csv	Sensors: Oculus Quest 2, Empatica E4	Understand participant's objective physical reactions to different interventions in the public space	OneDrive	The research team: the PI and three professors: dr. Yan Feng (TU Delft); dr. Jaime Soza Parra (UU); prof. Alexander Klippel (WUR)
Age range, occupation, household	.csv	Tablet, in the VR experiment lab (AMS Institute, Amsterdam) | I will recruit the participants through posters and professional and student networks.	To collect the age of the participants and occupation to make appropriate correlations	OneDrive	The research team: the PI and three professors: dr. Yan Feng (TU Delft); dr. Jaime Soza Parra (UU); prof. Alexander Klippel (WUR)
Questionnaires - transit habits	.csv	Tablet, in the VR experiment lab (AMS Institute, Amsterdam) Pre experiment questionnaire	To collect the ordinary behavior habits of participants and make appropriate correlations to their stress responses	OneDrive	The research team: the PI and three professors: dr. Yan Feng (TU Delft); dr. Jaime Soza Parra (UU); prof. Alexander Klippel (WUR)
Personal identifiable Information (PII) - Names, email	.csv	Provided by the participants themselves. It will not be published.	To contact them and schedule the experiment	OneDrive	Only the PI
Signed Inform Consent Form	.csv	Tablet, in the VR experiment lab (AMS Institute, Amsterdam)		OneDrive	Will be kept for 10 years to my TU Delft supervisor, dr. Yan Feng.

Appendix J: Acknowledgements

This thesis is a product of long and dedicated research, brainstorming, design, recruiting participants, running the experiment, analysis, and writing.

This thesis acknowledges the use of AI-based tools to help write, improve, and correct Python scripts and R code. The code has been used to analyze psychological and physiological data. In addition, AI has been used as a support tool for statistical analyses, quality checks, and code correctness. AI was used to support grammar and language tuning, and an AI-based engine to find similar VR studies for the discussion, but not as the primary literature review. The analytical decisions, data collection, interpretation of the results, and conclusions remain the intellectual product of the writer.

The TU Delft XR lab at the library supported the construction of the virtual reality using Unreal Engine, Rhino, and coding scripts, and provided access to laboratory resources. I would like to thank the XR lab at TU Delft, Arend-Jan Krooneman, and the team for supporting this research, helping with Unreal Engine, and providing the VR headsets and tools. The eXtended reality lab in TU Delft and Dr. Yan Feng for providing the Empatica wristband and phone. Layla Farmahini Farahani for supporting and brainstorming in several occasions about data collection methods.

I would like to thank my supervisors, Dr. Yan Feng, Prof. Alexander Klippel, and Dr. Jaime Soza Para, for their support, reviews, and brainstorming throughout this long thesis-writing process.

A special thank is dedicated to all forty-seven participant who voluntarily took part in the experiment and generously contributed their time and effort to this research.