Application of Virtual Reality in the study of Human Behavior in Fire

Pursuing realistic behavior in evacuation experiments

Arias, Silvia

2021

Document Version:
Publisher's PDF, also known as Version of record

Link to publication

Citation for published version (APA):
Arias, S. (2021). *Application of Virtual Reality in the study of Human Behavior in Fire: Pursuing realistic behavior in evacuation experiments*. Division of Fire Safety Engineering, Lund University.

Total number of authors:
1

Application of Virtual Reality in the study of Human Behavior in Fire

Pursuing realistic behavior in evacuation experiments

SILVIA ARIAS
DIVISION OF FIRE SAFETY ENGINEERING | FACULTY OF ENGINEERING | LUND UNIVERSITY

Printed by Media-Tryck, Lund 2021 NORDIC SWAN ECOLABEL 3041 0903

LUND UNIVERSITY

ISBN 978-91-7895-867-2
ISSN 1402-3504
ISRN LUTVDG/TVBB--1066--SE
Report 1066

Application of Virtual Reality in the study of Human Behavior in Fire

Pursuing realistic behavior in evacuation experiments

Silvia Arias

LUND UNIVERSITY

DOCTORAL DISSERTATION

by due permission of the Faculty of Engineering, Lund University, Sweden.
To be publicly defended at the Division of Fire Safety Engineering
on June 18th, 2021 at 13.00.

Faculty opponent
Dr. Peter Lawrence
University of Greenwich

Organization
LUND UNIVERSITY
Division of Fire Safety Engineering, Faculty of Engineering

Author
Silvia Arias

Document name
DOCTORAL DISSERTATION

Date of issue 25th of May 2021

Sponsoring organizations
Swedish Civil Contingencies Agency (MSB), European Organization for Nuclear Research (CERN), Craaford Foundation, Swedish Research Fire Board (Brandforsk)

Title and subtitle
Application of Virtual Reality in the study of Human Behavior in Fire – Pursuing realistic behavior in evacuation experiments

Abstract
Virtual Reality (VR) experiments are used to study human behavior in fire because they allow simulation of fire events with relatively low risks to the participants, while maintaining high levels of experimental control. Many studies have used VR experiments to explore aspects of the human response to fire threats, but VR experiments as a research method are yet to be subjected to a systematic process of validation. One way to validate VR experiments is to compare VR data to data obtained using other research methods, e.g., case studies, laboratory experiments, and field experiments. Five independent VR experiments were designed to collect data that could be then compared to data collected using other research methods. Both datasets, VR and physical, are then compared with each other to assess similarities and differences between them. Results show that participants in the VR experiments often acted like people did in the physical-world events. Moreover, Human Behavior in Fire theories that explain the behavior of victims in real fires were found to also explain the participants' behavior in the VR experiments. There were differences between VR and physical-world samples, which highlighted limitations of VR experiments or aspects about realism that need to be considered when designing VR experiments. Visual realism is not enough for participants to interpret a virtual fire emergency as a threat. Therefore, VR experiments need to induce participants to take the virtual fire event seriously. Social norms that are apply in physical world contexts may not emerge naturally in virtual environments, and measures should be taken to enhance behavioral realism in VR. These findings are a meaningful contribution to the development of the VR experiment method for collection of behavioral data.

Key words Virtual Reality, behavioral realism, fire evacuation, VR experiments

Classification system and/or index terms (if any)

Supplementary bibliographical information	Language
Report 1066 ISRN LUTVDG/TVBB–1066--SE	English

ISSN and key title	**ISBN**
1402-3504	978-91-7895-867-2 (print) 978-91-7895-868-9 (pdf)

Recipient's notes	**Number of pages**	Price
	178	-
	Security classification Open	

Signature

Date 2021-05-07

Application of Virtual Reality in the study of Human Behavior in Fire

Pursuing realistic behavior in evacuation experiments

Silvia Arias

LUND UNIVERSITY

Division of Fire Safety Engineering
Department of Building and Environmental Technology
Faculty of Engineering
Lund University

ISBN 978-91-7895-867-2 (print)
ISBN 978-91-7895-868-9 (pdf)

Printed in Sweden by Media-Tryck, Lund University
Lund 2021

Media-Tryck is a Nordic Swan Ecolabel certified provider of printed material. Read more about our environmental work at www.mediatryck.lu.se

MADE IN SWEDEN

Acknowledgements

I would like to express my gratitude to the many people who have supported me and made my research possible.

The experiments conducted during my PhD studies were made possible by the trust and support of the organizations who funded my research: the Swedish Civil Contingencies Agency (MSB), the European Organization for Nuclear Research (CERN), the Craaford Foundation and the Swedish Research Fire Board (Brandforsk).

I would like to thank my three supervisors, Daniel Nilsson, Enrico Ronchi and Håkan Frantzich, for the many discussions we had. They challenged me to keep moving forward, provided support, and were patient and caring throughout the years. Patrick van Hees, as head of the division, never missed an opportunity to listen and offer support. I am truly thankful to all of them.

Jonathan Wahlqvist spent a great deal of time patiently teaching me the skills I needed to create realistic virtual environments. We collaborated in the development of my VR experiments, and I am very grateful for his help.

I would like to thank Rita Fahy, who facilitated my long visit to the National Fire Protection Association. During my studies she acted as mentor, colleague, and informal supervisor.

Lastly, I would like to thank my family, for all their love and support: my dear husband, John; my parents, Carlos and Martha; and my siblings Ernesto, Celeste and María Inés.

Summary

In recent years, Virtual Reality (VR) has gained a foothold in the field of Human Behavior in Fire. VR experiments have been used to study human behavior in fire because they allow experiencing fire scenarios with relatively minor risks to participants and high levels of experimental control. While a large number of studies have used VR experiments to investigate different aspects of the human response to fire threats, their application for data collection has not yet gone through a systematic process of validation. One way to validate VR experiments for data collection is to compare the data they generated to data obtained from other sources (e.g., case studies, laboratory experiments, and field experiments).

Five independent VR experiments were designed to collect the same data collected using physical-world (non-VR) research methods. Both datasets, VR and physical, are then compared with each other to assess how similar they are. Each experiment was based on a different virtual environment (a two-story house, a hotel room, a high-rise hotel building, an underground particle accelerator, and a nightclub), and it was therefore possible to capture the behavior of participants in different virtual fire emergencies.

Results show that participants in these VR experiments often acted like people did in the physical-world environments the VR experiment represented. Each experiment exposed participants to a single virtual environment, in which participants exhibited different behavioral patterns. Moreover, Human Behavior in Fire theories that are commonly used to explain the behavior of victims in real fires were found to also explain the participants' behavior in the virtual context.

Participants were able to execute complex actions in VR, matching the behavior of people in the physical-world fire events. The differences between VR and physical-world samples pointed out limitations of VR experiments, or certain aspects about the realism of the virtual experience, that need to be taken into consideration when designing a VR experiment. For example, in some experiments it became clear that visual realism in a virtual environment is not enough for participants to interpret the fire emergency as a threat. Therefore, the scenarios in VR experiments, in addition to looking realistic, need to motivate participants take the fire event seriously. Moreover, the code of conduct that affects human behavior in physical-world environments may not emerge naturally in virtual environments, as it was shown by the rude or aggressive

behavior of participants towards non-player characters. This difference indicates that additional considerations need to be made to enforce social rules in virtual environments.

The contrast between the VR and the physical-world data showed the many ways the participants' perception of realism can be improved in modern virtual environments to enhance the behavioral realism of their VR experience. These findings are a meaningful contribution to advance the development of the VR experiment method for collection of behavioral data.

Table of Contents

1. Introduction

Human Behavior in Fire, which is a sub-field of fire safety engineering, is a research area focusing on people's response to fires and similar emergencies (Kuligowski, 2016). Even though Human Behavior in Fire is a relatively young field, long strides have been taken to reach its current state (Shields & Proulx, 2000). Initially, data was collected from witnesses' accounts through questionnaires and interviews (Bryan, 1983; Canter, Breaux, & Sime, 1980; Guylène Proulx & Fahy, 1997). However, experiments testing hypotheses and exploring causation are needed to ensure scientific rigor. Numerous experiments have been performed, testing hypotheses, measuring the impact of variables and drawing conclusions. Laboratory and field experiments have been widely used, providing invaluable knowledge for the improvement of the safety design of buildings (Kobes et al., 2010; Liao, Kemloh Wagoum, & Bode, 2017; Nilsson & Johansson, 2009).

However, the dangerous nature of fires limits the type of experiments that can be performed due to safety and ethical concerns. High temperatures and radiative heat flux can damage human tissue (Purser & McAllister, 2016). In addition, the inhalation of some combustion products can lead to long lasting health consequences, or even incapacitation and death if the dose or concentration is high enough (Purser, 2016). Therefore, the scientific impact of an experiment including fire and smoke may not compensate the risks for the participants.

Alternatives to real smoke have been used in experiments to minimize risks while still replicating low visibility conditions and irritant products in the smoke. Theatrical fog has been used to represent smoke (Latané & Darley, 1970), and mild irritants have been added to simulate the eye irritation caused by fire smoke (Fridolf, Ronchi, Nilsson, & Frantzich, 2013). These attempts to represent the effects of real smoke in experiments without its associated risks were able to reduce the visibility level as real smoke would do and cause some mild irritation as well. However, these focused in assessing an evacuation system rather than at reproducing the behavior of real fire victims, and it remains unclear whether these experiments replicate real victims' behavior.

Flames can also be hard to include in Human Behavior in Fire experiments. Intrinsically dangerous, the inclusion of large, uncontrolled open flames as those of a real fire with minimum risks to the participants is extremely hard. Therefore, there is a reduction in the ability to study the victims' attempts at fire suppression, compartmentation or rescue. These attempts, if unsuccessful, may lead to a delayed evacuation and to increased risks. Without being able to replicate dangerous conditions in a controlled experiment, the source of data remains observational.

In recent years, Virtual Reality (VR) has gained a foothold in the field of Human Behavior in Fire. VR experiments have a potential for studying human behavior in fire, since they allow simulation of fire and smoke with relatively minor risks to participants. Numerous studies have used VR experiments to investigate decision-making (Bode & Codling, 2018; Kinateder, Ronchi, Gromer, et al., 2014; Kinateder & Warren, 2016), way-finding (Ronchi, Kinateder, et al., 2015; Tang, Wu, & Lin, 2009), system design (Mossberg, Nilsson, & Wahlqvist, 2020; Ronchi & Nilsson, 2015), and evacuation behavior (Gamberini, Chittaro, Spagnolli, & Carlesso, 2015; Kinateder, Warren, & Schloss, 2019; Moussaid et al., 2016), along many other, advancing the development of VR as an experimental method.

In Human Behavior in Fire, VR experiments are a form of laboratory experiments (Kinateder, Ronchi, Nilsson, et al., 2014). As such, they allow for a relatively high level of experimental control. VR technology is able to recreate all sorts of environments: existing buildings (Andrée, Nilsson, & Eriksson, 2016), building projects in the design phase (Arias, Ronchi, Wahlqvist, La Mendola, & Rios, 2019), and theoretical, unlikely constructions (e.g., a never-ending corridors in which evacuation signage can be tested) (Troncoso, Nilsson, & Ronchi, 2015). Moreover, the risks for the participants are low, even when the VR experiments include virtual fire and smoke, making VR experiments a useful research method for Human Behavior in Fire experiments: the VR experiment method.

The VR experiment method is here defined as a research method consisting of the application of VR experiments for data collection in Human Behavior in Fire. The VR experiment method has unique disadvantages compared to other research methods. The disadvantages refer not only to the state of the art of the VR technology (such as computational power, ability to handle large groups of participants at the same time, area of coverage of the motion sensors, among other), but also to the nature of the VR technology: the virtual surroundings the participant sees are an illusion. As a modern version of the classic smoke and mirrors, VR is based in a system of screens and lenses, and what participants see while in VR, the *virtual environment*, is by no means reality. Although the virtual environment may look realistic and the objects in it may work like real objects would, participants are aware that they are in an artificial environment. While the same can be said about any laboratory experiments, there are unique intricacies in attempting to capture natural human behavior in an optical illusion.

Participants know the fire and smoke they encounter are not real, and therefore the risks associated with them are not the same. If participants expect the risks to be negligible, their behavior in the virtual environment may differ from that of fire victims.

The VR experiment method is introduced here not as a replacement for any specific research method, but rather a complementary one. The VR experiment method has advantages and disadvantages that need to be considered when selecting the most suitable research method for a given study. Nevertheless, the validation process of the VR experiment method has barely started, and more experiments contrasting VR and physical-world data are needed. Questions remain on how well VR experiments can elicit realistic behavior in participants and subsequently produce data at least as valid as that of other research methods.

Some experiments in Human Behavior in Fire do not hide the fact that they aim at studying an emergency or evacuation, and participants are told about it before they sign up (Jenssen et al., 2018; Kinateder et al., 2019; Ronchi, Nilsson, et al., 2015; Troncoso et al., 2015; Wetterberg, Ronchi, & Wahlqvist, 2020).Other experiments demand certain level of deception, and participants are not told in advance about any emergency taking place in the experiment, in order to collect data about their natural reaction to the situation they are exposed to. When information is to be concealed, it is important that the virtual environment both allows and motivates participants to behave realistically. Such a virtual environment is here defined as a *realistic virtual environment*.

A realistic virtual environment, therefore, has high levels of *behavioral realism*[1] (Steuer, 1992). Behavioral realism is defined as "the degree to which virtual humans and other objects within Immersive Virtual Environments behave as they would in the physical world" (Steuer, 1992). If the behavior of participants in a VR experiment is to be compared to that of people in the real world, the behavioral realism of the virtual environment cannot be overlooked.

An experiment conducted by Kisker, Gruber, and Schöne (2019) gives a good example of behavioral realism. In that experiment, an urban environment was simulated in VR. In it, a virtual high-rise building had a steel girder protruding from either the top of the building (treatment scenario) or ground level (control scenario). In the experimental room, a set of wooden planks was placed on the floor, matching the layout and the location of the virtual steel girder. Participants in each scenario were asked to walk on the steel girder in VR, which also meant them walking simultaneously on the wooden planks on the floor. The results

[1] It should not be confused with photorealism, which refers to the computer-rendered images being almost indistinguishable from a photograph. Photorealism is not a necessary or sufficient condition for a realistic virtual environment.

showed that participants in the virtual height condition (treatment) walked slower than those in the control group, and showed signs of anxious behavior, including a higher heartrate. Participants in the control group were overall more relaxed and did not show signs of insecurity while walking. This example shows how a virtual environment can motivate participants to behave according to their virtual surroundings, even when they know the risks they encounter there are inexistent in the physical environment they are in.

1.1. Identification of the problem

As mentioned before, the VR experiment method can be especially useful to study behavior in scenarios that are too dangerous to be reproduced in controlled conditions, such as laboratory or field experiments. In order to apply the VR experiment method for behavioral data collection in Human Behavior in Fire, its limitations need to be clearly understood. It is here suggested, that for a virtual environment to be realistic, the range of possible actions in the virtual environment needs to approach that in the real world. However, there is no clear advice or guidelines on how to produce a realistic virtual environment for VR experiments.

While many studies have used VR experiments to investigate human behavior in fire (Bode & Codling, 2018; Cosma, Ronchi, & Nilsson, 2016; Duarte, Rebelo, Teles, & Wogalter, 2014; Mossberg, Nilsson, & Wahlqvist, 2020; Moussaid et al., 2016; Shaw et al., 2019), the research rarely refers to the challenges of designing a virtual scenario that elicits realistic behavior. The projects usually have an engineering orientation, aiming to produce data to solve a specific problem in an efficient and cost-effective way. Consequently, the research objectives leave little room to address the peculiarities, challenges and pitfalls of creating realistic virtual environments for a VR experiment for behavioral data collection, as these are detached from the engineering objectives. VR-specific publications, on the other hand, are focused on more fundamental aspects of the VR technology than the specific issues of implementing VR in niche areas of application, such as Human Behavior in Fire. Therefore, Human Behavior in Fire researchers developing VR experiments may not be able to share the VR-related knowledge they gained through experience. Newcomers may need to learn on their own, with a high chance of repeating mistakes and learning the same lessons others already have. This inefficient use of resources has the additional disadvantage of impeding the refinement of the VR experiment method for behavioral data collection in Human Behavior in Fire.

Once the virtual environment is a fair representation of reality, it can be tested by comparing their results to the behavior observed in real fire incidents or in laboratory or field experiments. However, without guidance on what exactly a realistic virtual environment is like, validation of the VR experiment method is improbable. Before starting the discussion about validation of the VR experiment method, expertise in the generation of realistic virtual environments is needed. With this expertise, particularities, limitations, and possible ways to pursue behavioral realism in the VR experiment method can be brought up for discussion and assessment. The present work aims to start the discussion about behavioral realism in VR experiments for behavioral data collection in Human Behavior in Fire and to consider its effect in the data these experiments produce.

One way to assess how effective the VR experiment method is at reproducing human behavior is to compare the data collected from VR experiments to data obtained using other research methods. By comparing VR data to data from other sources widely accepted in the field of Human Behavior in Fire, it is possible to assess how well the virtual environment replicates the real conditions it intended to portray. Examples of these sources are case studies, laboratory experiments, field experiments, and fire drills (Kinateder, Ronchi, Nilsson, et al., 2014).

In the context of this research work, there are two types of data source: the data obtained using the VR experiment method (i.e., VR data), and the data obtained using any other research method (i.e., physical-world data). The term *physical-world* used throughout this research work refers to anything non-virtual. Sources of physical-world data are case studies, drills, traditional laboratory experiments and field experiments. In other words, physical-world is used here to refer to everything that is not VR. A laboratory experiment that takes place in a non-VR setup is here assumed to produce physical-world data that VR data can be compared to. Even if that physical-world data may not necessarily be valid outside the experimental conditions of the laboratory, as long as the VR data matches it, the VR experiment method will be deemed successful at replicating that specific set of physical-world data. With the definition of physical-world introduced, the distinction between laboratory experiments and field experiments does not play a major role in the context of this research work. Both laboratory experiments and field experiments belong in the physical-world realm, different from that of the VR experiments. Therefore, the term *physical experiment* is here introduced to refer to both laboratory experiments and field experiments that take place in the physical world, to distinguish them from VR experiments.

With those definitions in place, the attention can be focused on how to compare physical-world data and VR data: *what exactly should the VR data be validated against?* This question is certainly not unique to the VR experiment method, and it applies to any and all models humans can make. The straightforward answer is physical-world data, but this answer is hardly satisfactory. As mentioned before, experimental data does not necessarily represent the behavior of fire victims.

Moreover, the data that ideally could be collected from a single physical-world fire, may be *real* but it may not be observed in every possible real fire, since the many variables affecting the decision-making of each individual are not fully understood. Physical-world behavior may be a mathematical set of countless data-points of observed behaviors, none of them individually being an integral representation of the whole set. Since there is no absolute reality that can be measured objectively, any physical-world data that can be compared to VR data is here considered a good benchmark.

When comparing VR data to physical-world data, two aspects can be observed: whether participants in the VR experiment show similar behavioral patterns to those observed in physical-world events, and whether the VR data matches quantitatively the physical-world data. These two aspects indicate how similar the behavioral data produced in a VR experiment is to physical-world behavior.

The observed behavioral patterns are related to a qualitative assessment, and they refer to general Human Behavior in Fire concepts, such as perception of fire cues, decision-making, way-finding, suppression and compartmentation attempts, use of emergency exits, etc. If the data produced in a VR experiment is good, the behavior of participants should follow the same patterns observed in building occupants during a physical-world emergency. Those patterns would be present in any scenario, independently of the objective of the experiment. The second aspect, the match of VR data with physical-world data, refers to quantitative terms. These quantitative terms could be measured as proportion of participants doing the same actions as the people in a fire or in a physical experiment did. As an example, those proportions may refer to preference for the use of the available exits, or walking paths, or compliance with emergency signage, among others.

1.2. Objectives

The present research work will explore the suitability of VR experiments as a research method for collection of behavioral data in Human Behavior in Fire experiments, based on three objectives:

1. Investigate if the behavior of participants in VR experiments follows the same patterns reported in fire incidents.
2. Compare behavioral data obtained in VR experiments to that obtained from physical-world sources to assess differences between them.
3. Identify limitations of the VR experiment method and considerations to be made in the pursuit of behavioral realism when using the VR method for collection of behavioral data.

1.3. Publications

This research work is based on the four papers that have been submitted and accepted to relevant scientific journals detailed below. All papers have been fully peer-reviewed.

Paper I	Arias, S., Nilsson, D., & Wahlqvist, J. (2020). A virtual reality study of behavioral sequences in residential fires. *Fire Safety Journal*, https://doi.org/10.1016/j.firesaf.2020.103067
Paper II	Arias, S., Fahy, R., Ronchi, E., Nilsson, D., Frantzich, H., & Wahlqvist, J. (2019). Forensic virtual reality: Investigating individual behavior in the MGM Grand fire. *Fire Safety Journal*, 109, https://doi.org/10.1016/j.firesaf.2019.102861
Paper III	Arias, S., Mossberg, A., Nilsson, D., & Wahlqvist, J. (2020) A study on evacuation behavior in physical and Virtual Reality experiments. *Submitted for publication*
Paper IV	Arias, S, Wahlqvist, J, Nilsson, D, Ronchi, E, & Frantzich, H. (2020). Pursuing behavioral realism in Virtual Reality for fire evacuation research. *Fire and Materials*. 1–11. https://doi.org/10.1002/fam.2922

The author was involved with all phases of the four papers, from project conception to publication or presentation at a conference. Those phases are detailed as follows:

Project conception: consisted of the development of an idea for a project and the process of application for funding.

Experimental design: consisted of the development of the experimental plan and the design of the chosen scenario based on the data expected to be collected. It included the method to collect the data, the procedure of the experiment, testing the equipment, development of an associated questionnaire, application for ethical approval and information for participants.

Design of the virtual environment: consisted of the generation of the virtual environment where the experiment will take place. This phase consisted of the generation of the visual components and the programming of the interactions with the virtual environment.

a. Visual components: everything visible in the virtual environment (i.e., 3D models of the building, furnishings and surroundings, and lighting)

b. Programming of interactions: writing of scripts covering the possible ways the participant could interact with the virtual environment, such as opening and closing of doors, a working cellphone, a working TV with its remove controller, alarm triggers for the emergency simulated, fire growth or smoke fill-up as necessary, and generation of output files recording eye-tracking data or permanence in a given room.

Execution of the experiment: starting with the call for participants subsequent recruitment, this phase covered the implementation of the experimental procedure, training on use of the VR equipment, monitoring during the experiment, and debriefing session. In some cases, student helpers were hired for this phase, which required training and monitoring of the helpers.

Data analysis: it consisted of aggregation of the data produced, systematic examination of the results and questionnaire answers, the corresponding statistical testing and production of graphical representations.

Paper writing: this phase consisted of the production of scientific text that described the project thoroughly, presented the results obtained and highlighted relevant conclusions. The phase also included the submission of the paper to a peer-reviewed conference or to a relevant scientific journal, and the incorporation of changes reflecting the comments provided by the supervisors, coauthors and peer-reviewers when needed.

Presentation: related to the cases in which the paper was presented in a conference. This phase includes the participation as speaker in the conference, preparation of the presentation and its delivery in front of the audience, finalizing with question session afterwards.

Table 1 presents a detailed description of the contributions by the author in the four papers. The terms minor, medium and major refer to the level of the author's involvement. Minor involvement refers to up to 1/3 of the work; medium refers to between 1/3 and 2/3 of the work; and major refers to more than 2/3 of the work. *Paper IV* was based on previous experiments (some of them included in the other three papers). This paper was presented at the Interflam conference in 2019, and it was subsequently accepted for publication in the special issue of a scientific journal on the conference. *Paper I* was presented at the International Symposium of Fire Safety Science in 2021.

Table 1 – Level of the author's contribution in each phase the four papers in this research work

	Paper I	*Paper II*	*Paper III*	*Paper IV*
project conception	minor	minor	minor	medium
experimental design	minor	major	minor	n.a.
design of the virtual environment				
visual components	major	major	major	n.a.
programming of interactions	minor	minor	minor	n.a.
execution of the experiment	major	major	major	n.a.
data analysis	major	major	major	n.a.
paper writing	major	major	major	major
presentation	major.	n.a.	n.a.	major

1.3.1. Related publications

The following publications provide further information about some of the experiments discussed in the four papers.

Arias, S., La Mendola, S., Wahlqvist, J. Rios, O., Nilsson, D., Ronchi, E. (2019) Virtual Reality Evacuation Experiments on Way-Finding Systems for the Future Circular Collider. Fire Technology 55, 2319–2340 (2019). https://doi.org/10.1007/s10694-019-00868-y

Arias, S., Ronchi, E., Wahlqvist, J., Eriksson, J., & Nilsson, D. (2018). ForensicVR: Investigating human behaviour in fire with Virtual Reality. (LUTVDG/TVBB; No. 3218). Lund

Arias, S., Nilsson, D., Ronchi, E., Wahlqvist, J. (2017) Use of omnidirectional treadmill in virtual reality evacuation experiments, IAFSS 2017 poster 12th International Symposium of Fire Safety Science. Non peer-reviewed international conference poster.

2. Virtual Reality in context

This chapter presents the background knowledge needed to understand the terminology used throughout this research work with respect to VR. The chapter will not present an overarching description of VR, but rather it will refer only to the specific aspects of it that are relevant to this work. The following sections in this chapter will cover the concept of VR, which is here argued to be a two-fold concept: a technology and an experience. Then, each of those meanings will be detailed. VR as a technology will be presented in terms of the way it works, the most common VR equipment, and how locomotion can be implemented in VR. The section on VR as an experience will focus on the perception from the point of view of the user, including some terminology used to describe the feeling of being in VR. Lastly, the use of VR for research purposes will be summarized.

2.1. What is Virtual Reality?

Virtual Reality can be hard to define, as the term can be used to refer to a technology, or an experience created by said technology. Moreover, both the type of equipment used and the way the user operates in the virtual environment can produce substantially different experiences, even though the technology is the same. As a technology, for the purpose of this work, VR is a digital three-dimensional representation of an environment in which physical presence can be simulated. Artificial sensory stimuli such as sight, hearing, touch and smell can be added to simulate physical-world stimuli. The user can interact with the environment, reacting to it or altering it with their actions. The concept of VR as a technology focuses on technical aspects: images are produced, stimuli are simulated, and interactions are possible.

Virtual Reality as an experience, on the other hand, is the perception the user has of being in an environment created using VR technology. The user knows the environment they perceive is the product of a specific equipment, and it does not exist in the physical world. Nevertheless, the information the user received through their senses makes the virtual environment feel real. A VR experience can be conveyed by different types of equipment, but the equipment is only the medium, not the experience nor the technology. This notion is aligned with the ideas presented by Steuer (1992), who argued that the definition of VR should be based on the experience the technology provides rather than the equipment used to provide such experience: “A virtual reality is defined as a real or simulated environment in which a perceiver experiences telepresence”. This definition has the disadvantage of downplaying the role of the equipment used, as different types of equipment can offer largely different kinds of experience.

Evans (2019) presents a summary of the beginnings of VR as a technology, starting in the late XVIII century with Robert Baker’s Panorama. Craig, Sherman, and Will (2009) mention the Sensorama patented by Heilig (1962) and the 3D head-mounted display developed by Sutherland (1968) as two of the many primitive versions of modern VR equipment. None of those VR equipment is likely to convey the same kind of VR experience participants in the VR experiments run in the context of this research work had using a modern head-mounted display. Therefore, while VR as an experience is not necessarily defined by the equipment used, the equipment may be intrinsic to the kind of experience the user gets.

As explained, it can be difficult to separate the technological aspect from the experiential one. Therefore, the terms VR technology and VR experience will be used when deemed necessary to refer to one or another. Table 2 presents the definitions of four terms (VR, VR technology, VR experience and virtual environment) to be used throughout this work.

Table 2 - definitions of terms

Term	Meaning in the context of this work
VR	Virtual Reality – a simulated (or virtual) reality. This term can also be used as an adjective (e.g., VR experiments)
VR technology	the principles or systems that produce the virtual environment and deliver it to the user via a head-mounted display – this term takes the perspective of the equipment used to produce that simulated reality
VR experience	what the user lived or experienced while in VR – this term does not take into account the equipment, only the sensations it created in the user. It is based on a given virtual environment
virtual environment	the environment the user sees around them while in VR. A virtual environment is finite and it is carefully designed to produce a given VR experience through a set of events, virtual objects and features

2.1.1. Reality and virtuality

Virtual Reality (VR) should not be confused with Augmented Reality (AR). While VR and AR share some features, they belong in different sectors of the *virtuality continuum* (Milgram & Colquhoun Jr., 1999). The virtuality continuum consists of the spectrum between reality (a completely unmodelled environment) and virtuality (a completely modelled environment), as shown on Figure 1. Mixed reality is everything that exists between the two ends of the virtuality continuum. There is no clear boundary indicating where exactly the spectrum changes from predominantly virtuality to predominantly reality. Nevertheless, AR should be understood as a mix of reality and virtuality in which most components are physical. There are several examples of experiments performed using AR for fire evacuation and training (Catal, Akbulut, Tunali, Ulug, & Ozturk, 2020; Saunders et al., 2018).

Figure 1 - Adaptation of the virtuality continuum presented by Milgram and Colquhoun Jr. (1999), showing schematically where AR and AV are within the mixed reality spectrum.

The VR experiment method and the VR experiments described in this research work belongs in the virtuality domain (see Figure 1). The incorporation of physical components will be discussed in other sections of this research work, which will mean a minor incursion in the Augmented Virtuality (AV) region. AV is analogous to AR, although on the other side of the virtuality continuum, and it consists of the addition of physical elements to a virtual context. An example of AV can be seen in the VR experiment run by Månsson (2018), in which the participant in the VR experiment had to pick up a physical-world fire extinguisher in order to operate the virtual one in the experiment. Blomander (2020) conducted another experiment in AV, in which the added physical element was thermal radiation. Radiative heat panels controlled by a computer allowed to mimic thermal radiation from the virtual thick layer of smoke in the virtual environment. While some of the content presented in this research work may also apply to AR experiments and furthermore to AV experiments, the work presented in the following chapters refers solely to VR.

2.2. VR as a technology

In this section, VR will be presented as the technology used to create a simulated reality and allow the user to feel like they are in it. The focus in this section is put on the way the technology works, and some common VR equipment. Locomotion in VR will be described, as it is also closely related to the VR equipment used.

2.2.1. VR equipment

Modern VR technology is based on mimicking stereoscopic vision. The images the user sees while in a virtual environment are the result of a deliberate illusion.

Two nearly identical images are presented, each to one eye in the right way to be perceived as one at a given distance (Wann, Rushton, & Mon-Williams, 1995). The brain is tricked into processing those two independent images as one, and the user perceives depth in the virtual environment. A set of motion sensors (e.g., accelerometers, gyroscopes) identify changes in the user's standpoint and adapt the images to be displayed accordingly. Different types of VR equipment use different methods to mimic stereoscopic vision, but the result is the same.

As new VR equipment are being introduced to the market at a rapid pace, it is not possible to cover all available and coming innovations. This section gives a brief overview of two kinds of VR equipment to give a basic idea of the most common types. The VR equipment presented here will be later brought up in the description of the experiments and the data produced.

The head-mounted device (HMD) is currently the most popular VR equipment. It consists of a sort of goggles to be strapped to the users' faces. The goggles include two screens, each placed in front of each of the user's eye, in which the corresponding images are displayed. The HMD is connected to a computer (tethered HMD), which renders the images seen on the screens. Some HMD have a computer integrated in the goggles and do not need to be connected to an external computer (untethered devices). Untethered devices may also be based on a smartphone, in which case the smartphone's screen is parted in two to present each eye the corresponding half through a set of lenses.

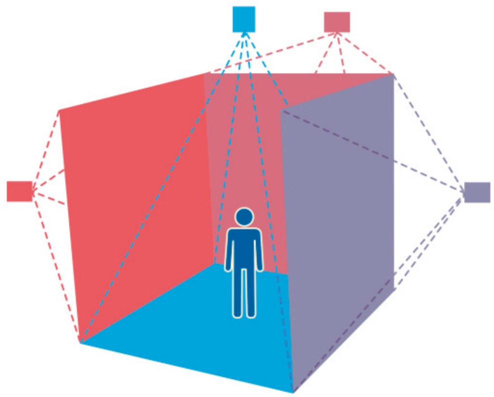

Figure 2 - Schematic diagram of a CAVE consisting of a system of four screens and projectors

An alternative to the HMD is the Cave Automatic Virtual Environment (CAVE). The CAVE consists of a set of projectors and large screens (about the size of a wall in a room), connected to a computer, which renders the VR images. Instead of goggles with screens, the CAVE has the screens set up roughly resembling a room in which the user stands. The CAVE surrounds the user, as illustrated on Figure 2, actively placing them within the virtual environment. A CAVE can consist of several screens surrounding the user, but likely the user could always see parts of the physical space (e.g., the ceiling, the motion sensors, joints between the screens, etc.).

Other relevant VR equipment in the context of this work are the hand-controllers. Hand-controllers are commonly used to interact with the virtual environment. Some hand-controllers can be tracked by motion sensors and mimic some functions of the human hand (e.g., grabbing objects and operating them). Additionally, hand-controllers can also be used for locomotion in VR.

2.2.2. Locomotion in VR

Navigation is the act of moving within the virtual environment. Navigation is achieved through different types of locomotion in VR. In the context of this work, the term navigation is used as an umbrella term to refer to the participant moving in the virtual environment, while locomotion is used to refer to the specific technique used to achieve that movement, in terms of equipment and commands needed to be executed by the participant. Four techniques of VR locomotion are prevalent: teleportation-based, controller-based, motion-based, and room scale-based. These techniques are summarized and presented in this sub-section, following the typology proposed by Boletsis (2017) illustrated on Figure 3. According to that typology, motion can be continuous or non-continuous.

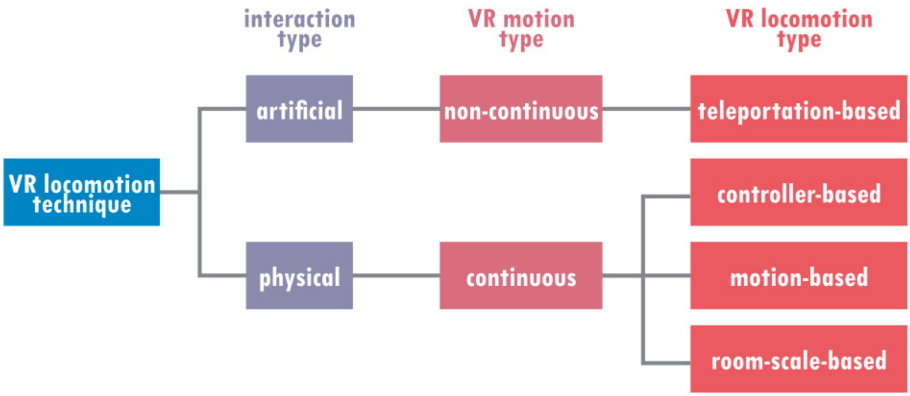

Figure 3 - VR locomotion typology proposed by Boletsis (2017). Figure adapted from Boletsis (2017)

In non-continuous locomotion, the user changes position abruptly, from point A to point B, without any intermediate steps. *Teleportation-based motion* is the only non-continuous motion. It consists of the user indicating a location within the virtual environment and being instantly placed in that location upon command. There is no journey between the two points, there is no movement speed, and there is no obstacle avoidance. The user cannot stop midway and change their destination. Only once the user reached the initial destination they can reassess and choose an alternative destination.

Continuous motion is a more realistic representation of physical-world movement. The user starts in one location and chooses where to go, but much like walking, running or driving a vehicle, the journey between the two points is part of the experience. This type of VR motion allows for changes of speed along the way, and obstacle avoidance may be needed to reach the desired location. Boletsis (2017) identified three types of locomotion within the continuous motion, which are: *controller-based*, *motion-based* and *room scale-based*.

Controller-based locomotion is considered an artificial interaction type by Boletsis (2017) because it relies on the use of a form of hand controller and therefore low intensity of physical activities. A joystick is a simple example of controller-based locomotion, but there are also systems that relate the movement to the direction the user is facing and turning the body to some degree may also be part of this type of locomotion. Moreover, some HMD may include their own set of hand-controllers that in addition to the locomotion functions, can be used to interact with the environment (e.g., pick up objects, open doors, etc.).

In the case of motion-based locomotion, the user navigates the virtual environment by making certain movements or physical activities of moderate or high intensity (Boletsis, 2017), like swinging their arms, kicking their feet, or by means of an omnidirectional treadmill. The movement speed in the virtual environment may be paired with the speed of the motions made in the physical world to give the user control on their movement speed.

Lastly, room scale-based locomotion allows the user to move freely in the physical environment as the virtual environment fits within it. In this type of locomotion, the movement of the human body in the physical environment is directly translated to the virtual environment (Boletsis, 2017). Walking, running, jumping, or other movements in the physical room are reflected in the virtual environment, making it possible for the user to move around as in reality.

Each VR locomotion type has advantages and disadvantages that may be more or less relevant depending on the scenario under study (Boletsis & Cedergren, 2019). Factors like demographics (e.g., elderly people or people with movement impairments may struggle to use an omnidirectional treadmill), available space (e.g., room-scale motion may not be possible if the physical space is too small), and objective of the study (e.g., being able to run in VR may be irrelevant in a study about visibility of emergency signage within smoke) need to be considered.

2.3. VR as an experience

When VR is defined as an experience, the focus is placed in the perspective of the user. The VR equipment used may have a major influence in the kind of experience the user gets, because of it being the medium through which the user accesses the VR experience. Nevertheless, the user's perceptions (i.e., how they felt, how they acted and why) are of high relevance when realism is to be considered.

2.3.1. The virtual environment

The virtual environment was defined in Table 2 as "the environment the user sees while in VR". In the same way an environment exists in reality, a virtual environment exists in Virtual Reality. Some environments in reality are natural (like a forest, a mountain range, a desert), and some are manmade (like a neighborhood, a crop field, a building). Virtual environments can replicate natural and manmade environments.

The virtual environment can be designed to simulate any environment, either existing or not: an urban environment (Kisker et al., 2019), a particle accelerator (Arias et al., 2019), a beach (Blum, Rockstroh, & Göritz, 2019), a forest (Browning, Mimnaugh, van Riper, Laurent, & LaValle, 2020), an endless corridor (Blomander, 2020), ancient Pompeii (Demetrescu, Ferdani, Dell'Unto, Leander Touati, & Lindgren, 2016), just to cite some.

The visual realism (i.e., how realistic the virtual environment looks), in terms of objects' appearance, textures, lighting, is only one aspect of the virtual environment. Interaction with elements of the virtual environment is an important component of the VR experience. Hand-controllers or any alternative hand-tracking device can allow the user to handle objects, operate doors, push buttons, etc., as mentioned before. Events can be added to the virtual environment to make the experience even more interactive. These events are highly dependent on the VR experience the designer of the virtual environment aims for. Examples of events relevant in the context of this work are the triggering of a smoke alarm, smoke starting to enter a room, a crowd evacuating the premises.

2.3.2. Single-player and multi-player experiences

The VR experience can be designed to be either single-player or multi-player. A single-player virtual environment can only take a single user (or player) at a time. Two or more users can be in a multi-player virtual environment at the same time and interact with it and with each other. Each user can be an independent actor within multi-player virtual environment. Both single-player and multi-player virtual environments can include computer-generated characters. They are the so-called *non-player characters* to differentiate them from the user-controlled counterparts because they are not operated by a user but by the computer, often by direct scripting of sequences or by computer algorithms of varying degrees of complexity. Their interactions with the user, if any, are rule based, usually programmed in advance, although artificial intelligence could also be used (Sharma et al., 2019). The non-player characters can be props (like pedestrians walking on the street, or customers in a café) that do not interfere with the user's experience but make the scene more realistic, or can be an active part of it playing a role or engaging with the user in some way.

2.3.3. Point of view

The user can be given a first-person perspective or a third-person perspective. When the user is given the first-person perspective, they see their surroundings in the virtual environment from the point of view of their eyes, in the same way as they do in the physical world. When the third-person perspective is given, the user can see the character they embody from a given distance, as a witness of what the character does, even though the user is in control of it. A first-person perspective is preferable for a realistic VR experience, as it resembles the way humans see their surroundings in the physical world.

2.3.4. The VR experience

Once the VR experience starts, the user is fully aware that what they see in the virtual environment is not real. Nevertheless, they may still act as they would do in the physical-world. Flinching, squinting and general obstacle avoidance occur on a regular basis, even if the virtual environment looks highly cartoonish. In fact, photorealism may not be the strictly required in order to get a realistic experience from it (Hoorn, Konijn, & van der Veer, 2003), or to have the feeling of being physically present in the virtual environment (Wright & van Waveren, 2014; Zibrek, Martin, & McDonnell, 2019), which is a cornerstone of the VR experience. Furthermore, achieving photorealism in VR may take a while: some VR experts have assessed that VR equipment needs to increase its power by a factor of 200 to achieve photorealism (MCV Staff, 2015), with others giving a timeframe of a couple of decades until then (Coleman, 2017).

Other concepts are used to describe the realism of the VR experience: presence, immersion, emotional response, engagement, and interactivity. These concepts are presented here in a rather simplified way, as a basic introduction to readers unfamiliar with them.

Presence is a recurrent concept when discussing the experience in a virtual environment. The term, however, seems to lack consensus in its definition and it may be used by different groups to refer to different concepts (Slater, 2003). Because there may not be a *right* definition of presence, as Slater (2003) states, it is important to clarify what is meant when using the term. In the context of this research work, the simple definition given by Slater, Usoh, and Steed (1995) is adopted: "presence is the psychological sense of 'being there' in the environment: it is an emergent property based on the immersive base given by the technology". A user feeling present has the sensation of being in the virtual environment in the same way they may feel present in a physical-world environment.

The term *immersion* is it is usually mixed up with presence. Immersion can be considered inherent to the VR equipment used. Slater (2003) proposes understanding immersion as follows: "let's reserve the term 'immersion' to stand simply for what the technology delivers from an objective point of view. The more that a system delivers displays (in all sensory modalities) and tracking that preserves fidelity in relation to their equivalent real-world sensory modalities, the more that it is 'immersive'". It can be argued that immersion and presence are independent from each other, but as Slater proposes, they are probably related (Slater, 2003). In the context of this work, immersion is an attribute of the VR equipment, while presence is the perception of the user.

Other, less frequent terms describe other aspects of the virtual experience. *Emotional response* is one of them. Users can have emotional responses to the events in the virtual environment, such as anxiety (Andreatta et al., 2020), stress (Chittaro & Zangrando, 2010), fear (Gromer, Reinke, Christner, & Pauli, 2019), empathy (Schutte & Stilinović, 2017), among other. *Engagement* (often referred to as *involvement*) refers to how much attention the user dedicates to the virtual surroundings (Gutierrez-Maldonado, Gutierrez-Martinez, Loreto, Peñaloza, & Nieto, 2010). Low engagement indicates that the user does not feel prompted to act or react to the events in the virtual environment. *Interactivity* has been defined by Steuer (1992) as "the extent to which users can participate in modifying the form and content of a mediated environment in real time". Therefore, interactivity refers to how much the user can influence the virtual environment, and it has been identified as a key characteristic of the VR experience (Mütterlein, 2018).

Even when experiencing a high level of presence, engagement and interactivity, it is unlikely that the user gets to completely forget about their physical surroundings while in VR, but at least their attention is displaced from the physical room to the virtual environment. A major part of the VR experience relies on the user playing along with the illusion. The user knows what they are seeing is not real, and them acting according to the virtual environment is partially their will. It is possible for them to completely ignore projectiles flying towards them. They may flinch, they may blink as part of a defensive reflex (Fossataro, Tieri, Grollero, Bruno, & Garbarini, 2020), but knowing nothing will hit them in reality, they may choose to ignore the projectiles despite the reflex reaction. Nevertheless, the virtual environment can be inciting enough that users need determination and active efforts to ignore it. The level of motivation can be an attribute of the virtual environment, and it is related to the concept of behavioral realism presented in Chapter 1.

These concepts refer to sensations or feelings the user experiences that can be difficult to measure objectively. However, as seen in the heart rate measurements in the experiment by Kisker et al. (2019), VR can elicit measurable physiological reactions too, even unintended ones. More than accelerated heartrate, perspiration and other indicators of stress, users may experience VR sickness, which symptoms are similar to those of motion-sickness (Gavgani, Walker, Hodgson, & Nalivaiko, 2018). The symptoms can range from being a mere nuisance to being intolerable for the user. There is a vast body of research on causes and mitigation measures (Chardonnet, Mirzaei, & Mérienne, 2017; Fernandes & Feiner, 2016; Guna et al., 2019; Munafo, Diedrick, & Stoffregen, 2017; Rebenitsch & Owen, 2016; Saredakis et al., 2020; Weech, Varghese, & Barnett-Cowan, 2018; Yildirim, 2020). The incidence of VR sickness have been found between 15 and 100% of participants in different studies (Chang, Pan, Tseng, & Stoffregen, 2012). The severity, however, may vary. Users may experience no symptoms or only mild ones, some may need a break from the VR experience, and in severe cases they may not be willing to resume. The symptoms can last from a couple of minutes after ending the VR experience to several hours.

2.3.5. Assessing the VR experience

Behavioral realism was defined before as “the degree to which virtual humans and other objects within Immersive Virtual Environments behave as they would in the physical world” (Blascovich, Beall, Swinth, Hoyt, & Bailenson, 2002). This definition is based on a comparison between the behavior observed in VR and the expected behavior in a physical-world version of the virtual environment. Behavioral realism is the most important concept in the context of this work.

It should be noticed that the definition of behavioral realism does not refer to the sense of presence. Presence may play a role on behavioral realism, but by definition it is not a necessary attribute of behavioral realism. With that respect, emotional response, engagement and even interactivity may be better indicators of behavioral realism than presence. A user may feel present in an environment but with low engagement, their behavior would not necessarily be the same it would be in a physical-world environment. It is possible, however, that the emphasis put on presence in other scientific publications is due to them using a different definition of presence. As mentioned before, there is no consensus. If their definition of presence assumes that feeling present implies a corresponding emotional response and high engagement, then presence is the sole most important concept to consider.

2.4. Applying VR in research

In a VR experiment, the participant is the user. The virtual environment is designed to include the experimental conditions in the VR experience. A single virtual environment can be used in different *scenarios*, each scenario presenting a variation of the original virtual environment. The distinction between virtual environment and scenario is here made because scenarios are especially relevant in the context of VR experiments, as one can be made as a control, and another one as a treatment. Differentiating the two concepts is also relevant, as some parts of this research work may refer to aspects of the virtual environment and some refer to a specific scenario of those based in the same virtual environment. Therefore, it is expected that anything said about virtual environments is valid in all corresponding scenarios, while what is said about a scenario does not necessarily apply to other scenarios.

The main advantage of VR may be the fact that it allows researchers to run scenarios that may be unfeasible in real life. Such scenarios may be too expensive (e.g., shutting down the Large Hadron Collider for a week (Arias et al., 2019)), or too risky (e.g., replicating a fatal fire in a nightclub (Arias, Ronchi, Wahlqvist, Eriksson, & Nilsson, 2018)), or the case could also be that the most suitable scenario is not necessarily realistic (e.g., an experiment to test the concept of homuncular flexibility (Stevenson Won, Bailenson, Lee, & Lanier, 2015)). Additionally, the risks associated with the scenario may be reduced in VR. For example, exposing participants to smoke and fire conditions may be too risky. Alternatives to physical smoke have been used in laboratory experiments (Fridolf et al., 2013; Latané & Darley, 1968). Maintaining the exact same smoke conditions for each run of the experiment may be a challenge. Flames are even harder to fake than smoke, and physical flames can very easily get out of control, making exposure of participants to them very risky. Exposing a participant to virtual fire and virtual smoke, however, is relatively simple. The computer-generated smoke and flames are controlled by algorithms designed to make the fire grow and the smoke spread according to the needs of the simulated scenario. Similarly to the theatrical smoke, the virtual smoke (called *physically based smoke*) lacks the smell, the irritants and the toxicity of physical smoke, but it can have the same light absorption and light scattering properties (Wahlqvist & van Hees, 2018). The ability of computer algorithms to replicate physics models makes physically based smoke a very useful tool to study visibility within smoke. Flames can be programmed to grow, spread or extinguish as needed, with no added risks to the participant. In addition to the visual representation of fire and smoke, radiative heat panels can be applied to complement the experience with thermal radiation (Blomander, 2020; Lawson et al., 2019).

Naturally, there are disadvantages that are specific to the VR experiment method. Some VR scenarios can only be run using a specific kind or even brand of equipment, becoming unusable once the specific equipment is discontinued or obsolete. Moreover, in most cases, commercially available HMD often reduces the field of view to roughly 100 degrees in the horizontal. VR sickness can affect some participants, with symptoms strong enough for them to terminate the experiment abruptly at any point, even before any data is collected. Collecting walking speeds could be difficult if the type of locomotion does not allow the user to walk freely, without being afraid of hitting boundaries in the physical surroundings or damaging the equipment. Dexterity of the participants using the equipment may affect the participants' performance, especially when the sample includes the elderly (Cook, Dissanayake, & Kaur, 2019; Ijsselsteijn, Nap, Poels, & De Kort, 2007).

3. Method

To achieve the objectives presented in Chapter 1, a research strategy was developed to compare data obtained from VR experiments (VR data) to physical-world data. The thorough description and analysis of different research methods and data collection techniques made by Nilsson (2009) was used to define research methods in the context of this research work. Following the definitions given by Nilsson (2009), two research methods are identified: case studies and experiments. The latter consists of laboratory experiments or field experiments. Any of these research methods could be used as source of physical-world data to be compared to VR data.

The research strategy consists of collecting behavioral data in a VR experiment based on either a well-documented case study or a physical experiment. The data obtained from the VR experiment will then be compared to that from the case study or the physical experiment.

Not any physical-world event (either a case study or a physical experiment) can easily be reproduced in a VR experiment. Three factors were identified, that determine whether a physical-world event could be reproduced in a virtual environment: availability of the physical-world data, identification of one or more behaviors of interest, and reproducibility of the chosen event in VR. These factors were derived from numerous attempts at recreating certain physical-world events for the VR experiments run in the context of this work. The reasons different physical-world events were not suitable to be reproduced in VR usually were of three different kinds, hence the three factors. Each will be described in detail.

Availability of physical-world data refers to whether enough data exists and is accessible. For example, investigation reports of past fire incidents (which are suitable for case studies) do not always publish the information they gathered in a detailed way or summarize similar witnesses' accounts into a single description. The high stress levels during the event, the lack of documentation like videos or pictures during the fire, the disparity between witnesses' accounts, and the missing pieces of information make it difficult for the designer of the virtual environment to ensure it presents the same conditions victims were subjected to. If the virtual environment does not replicate the same conditions, it is not reasonable to expect the same behavior. Therefore, case studies can be difficult to use as sources for physical-world data to be reproduced in a VR experiment.

Nevertheless, in many cases there is enough level of detail for investigation reports to be used.

Physical experiments, on the other hand, usually offer enough data to be replicated. Being those experiments designed by researchers with similar interest for meticulousness and experimental control of the chosen event, there are detailed descriptions and dimensions of the experimental setup, description of the ways the data was collected, the instruments used and their precision. The raw data is not always available, or may be incomplete, as it is presented in aggregated plots or averages. If the raw data is available, either in the publication or by its authors, it can be compared to the results from the VR experiment.

Behaviors of interest refer to the specific behavioral pattern or measurable dataset that can be collected from a given event. A behavior of interest could relate to decision-making (e.g., pre-evacuation time), route choice (e.g., which means of egress are used), actions performed (e.g., pre-evacuation activities), search for cues, compliance with evacuation signage, among others. A behavior of interest needs to be unambiguous to avoid misinterpretations of the data, and it needs to be measurable in some way (e.g., how many times an action was performed, or when did the participant start their evacuation, did they walk or run).

As mentioned before, behavioral patterns refer to attitudes and behaviors regularly observed in fire and evacuation events, as described by the relevant Human Behavior in Fire theories. These theories such as Behavioral Sequences (Canter et al., 1980), Theory of Affiliation (Sime, 1985), Role-rule Model (Tong & Canter, 1985), Social Influence (Deutsch & Gerard, 1955), and Risk Perception (Tancogne-Dejean & Laclémence, 2016) refer to behaviors that are commonly observed independently of the event itself. If the same behavioral patterns are observed in a VR experiment, it can be concluded that the VR data reflects reality to some extent. The behavioral patterns need to be easily identifiable during the VR experiment. For example, people show a tendency to evacuate the building through the everyday entrance/exit (Sime, 1985). In a VR experiment that offers several evacuation routes, this tendency, which is often called affiliation, should be easy to identify. If it is identified, it can be concluded that participants in the VR experiment showed the behavioral pattern just as building occupants would in a physical-world event.

Behavioral data, on the other hand, can be measured or counted in some way (e.g., frequency: how many participants did a given action; time: how long did it take for each participant to leave the building, etc.). As an example, determining when a participant decided to evacuate can be troublesome, since it is not possible to pinpoint the exact time a decision was made. Instead, the threshold could be the time the participant left a given room, which gives a clear-cut definition of when the measurement is made. This highlights the importance of measuring the data in an analogous way in both datasets to be comparable.

Replicability in VR refers to how suitable the event is to be replicated in VR. Even if enough data is available, some crucial conditions in an event may not be able to be replicated in VR. For example, high-density crowding conditions are hard to replicate in VR because of the difficulty of providing realistic crowd pressure based on the participant's movements in the experimental room. Overcrowding and even blockage at the exits have been critical in some fires, hindering the evacuation and forcing building occupants to change their chosen evacuation route (Comeau & Duval, 2000; Grosshandler, Bryner, Madrzykowski, & Kuntz, 2005). Dark environments are also hard to replicate in VR. Without the faculty of feeling their surroundings, a dark environment is unlikely to give the participant enough information about what is going on, or options on how to respond. Fire events may lead to darkness once the power supply is affected by the fire, making the event difficult to reproduce in VR. The experiment by Nilsson, Fridolf, and Frantzich (2012) can be used as an example. In that experiment, participants had to walk in a dark road tunnel filled with artificial smoke. Due to the low visibility conditions, many participants walked with their arms stretched in front of them, or put a hand on the tunnel wall as they walked. While participants in a VR experiment can do the same gestures, the lack of physical surroundings will prevent them to get any information. A VR experiment may fail to replicate the behavior of participants finding directions by touching the tunnel wall while walking, given the lack of sense of touch.

Replicability can also consider the differences between the level of risk in a fire and in a virtual one. In some fire events, victims are exposed to dramatic situations in which they need to make a difficult choice. Facing serious threats of injury by fire and smoke, building occupants in fires sometimes resort to dangerous actions that may mean the difference between life and death. An example of this is the patrons jumping out of the windows in the Gothenburg nightclub fire in 1998. Faced with dire options, many patrons resorted to jumping out the windows, taking a fall of around 6 m, which naturally resulted in injuries (Comeau & Duval, 2000). In a virtual environment, participants know the fire will not burn them, the smoke will not affect their breathing, and even if they need to jump out of a window in the virtual environment, they will not get hurt as they will not be falling in the physical world. When the participants' well-being is not on the line, there are no risks for them to consider as carefully as they people do in a physical-world case. The difference in the risk assessment in VR and in reality, may lead to the VR data overestimating the propensity to take a risk in physical-world case.

Furthermore, the validity of the VR data to be collected needs to be considered. For example, an experiment on walking speeds during an evacuation can easily be replicated in VR. However, the VR locomotion type chosen for the VR experiment will affect the validity of the walking speed data collected. Teleportation-based, controller-based and motion-based locomotion types do not allow natural walking or running, and the data they can produce on walking speeds may not be representative of physical-world data from a quantitative point of view. Room-scale-based locomotion could be suitable for collecting this data, if the experimental room is large enough to accommodate the area the participant will walk on and the VR equipment can cover it. Therefore, the replicability in VR should also reflect on the validity of the data to be collected.

3.1. Selection of fire events

As a starting point, the table on comparison of research methods presented by Kinateder, Ronchi, Nilsson, et al. (2014) was used to identify typical research methods that could supply physical-world data. This table, reproduced here in Table 3, presents a summarized comparison between VR experiments and other research methods in terms of setting, experimental control, ecological validity, replicability, among others. Table 3 adds drills to the research methods presented by Nilsson (2009).

As shown in Table 3, the VR experiment method can be contrasted with the other research methods by comparing specific aspects of them. It is not surprising that the VR experiment method is more similar to the classic laboratory experiments than any other method, as VR experiments are a form of laboratory experiments. Having identified the different research methods that could be used as a source of physical-world data, the next step was to determine which fire events could be reproduced. In each case, a specific behavioral pattern or a measurable set of data would be used as behavior of interest for comparison with the VR data.

The selection process for a physical-world data source started by suggesting a specific fire event with extensive documentation. The documentation, usually in the form of a fire investigation report and additional research outputs, was crucial to have a picture of the event and replicate it in the virtual environment. In some cases, it was necessary to contact the researchers to ask for their raw data. That documentation was used to identify behaviors of interest in it. Once the behaviors of interest were chosen, the replicability in VR was analyzed. The analysis was based on what features would be needed in the virtual environment to collect the data on the behavior of interest in terms of programming of the virtual environment and whether additional equipment was needed. An assessment was made on how well the fire event’s conditions could be incorporated in the virtual environment, and whether additional equipment could be obtained to improve either the experience (e.g., a wireless solution) or the measurement of certain behaviors of interest (e.g., use of emergency signage could be measured with eye-tracking devices). If the physical-world event satisfied all requirements, it was selected as an object of study and it was reproduced in VR. Figure 4 presents a schematic diagram of the selection process.

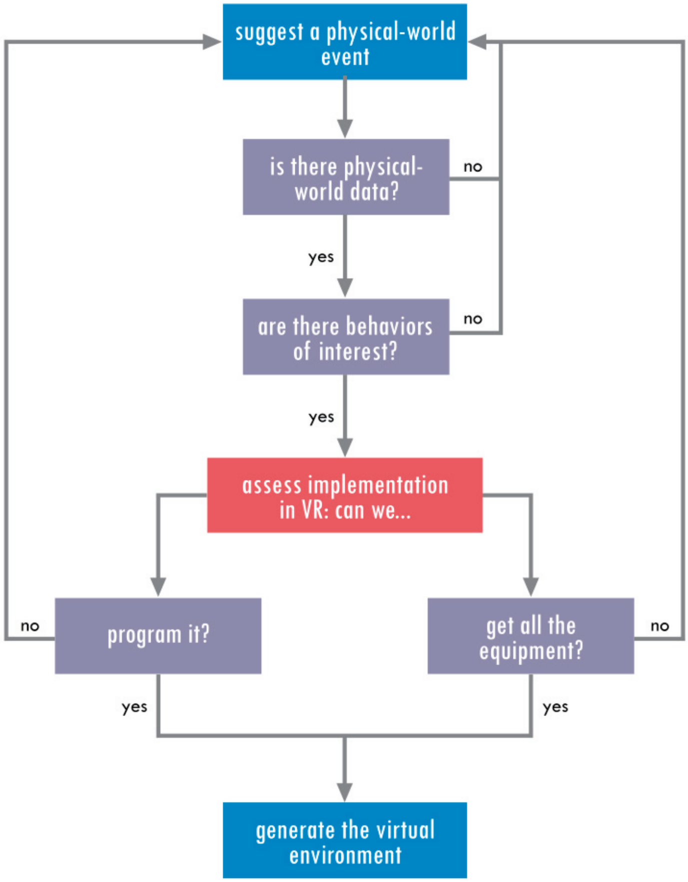

Figure 4 - Diagram of the process of selection of fire events to be reproduced in VR experiments

Table 3 – Reproduction of "Table I. Comparison of research methods" presented by Kinateder, Ronchi, Nilsson, et al. (2014). Note that the setting is classified as laboratory or real-world. With the exception of the VR experiment listed on the table, both laboratory and real-world settings are in this research work part of the physical-world.

	Hypothetical study	"Classical" lab experiment	VR experiment	Field studies	Drills	Case studies
setting	laboratory	laboratory	laboratory	real-world	real-world	real-world
experimental control	yes	yes (less than in VR)	yes	limited	no	no
type of data	subjective (statements from participants or experts)	subjective, objective (behavior & psycho-physiology)	subjective, objective (behavior & psycho-physiology)	subjective, objective (behavior)	subjective, objective (behavior)	subjective, objective (behavior)
possibility of use of stressors	no (only hypothetical)	limited	limited	limited	limited	yes
ecological validity	low	medium	medium	medium	high, if unannounced; limited if announced	high
possibility of adjusting experimental setting	yes	yes	yes	limited	no	no
possibliity of exact replication	yes	yes	yes	limited	no	no
time and cost intensity for data collection	very low	low	low	high	medium	-
automatic data collection possible	yes	yes	yes	limited	limited	no

3.1.1. Behaviors of interest

As mentioned before, behaviors of interest, replicability in VR and availability of the data were the factors considered for selecting a fire event to be replicated in a VR experiment. While the three factors are here considered equally important for selecting the fire event, the behaviors of interest are the core of the experiment to be designed, as they define the hypothesis and the results of the experiment.

The behaviors of interest refer to measurable behavior of fire victims that could be replicated in a virtual environment. All actions that victims performed in a fire are possible behaviors of interest, but only those that could be reproduced or recorded in VR were included. A good example are the events in the Station Nightclub fire, in which on top of poor visibility conditions, there were blockages at some of the few available exits (Grosshandler et al., 2005). Since low visibility and lack of sense of touch reduce dramatically the amount of input a participant would receive in a virtual representation of the Station Nightclub fire, this event could not produce comparable behavior to that observed in the fire event.

Several behaviors of interest were identified in the context of this research work, such as actions performed, walking paths and visual cues. These behaviors of interest will be discussed in the following subsections.

3.1.1.1. Actions performed

The actions performed by individuals in both the physical-world and the virtual samples are a convenient behavior of interest. They are relevant when studying the decision-making process. Decision-making is related to the assessment made by the individual about the cues they are aware of, their own assessment of the risk the situation entails, and it finalizes with the course of action they consider most appropriate given the circumstances. The decision-making cannot be studied in its pure form, as it can only be observed indirectly in the actions performed. In some cases, the actions performed can be very specific (e.g., calling the fire and rescue services), while in other cases several actions can be due to the same decision (e.g., attempting fire suppression by using an extinguisher, water, suffocating the flames, or other any other means). More than the specific action on itself (using an extinguisher or using water), the outcome (attempting fire suppression) is what matters when studying the decision-making process.

Comparing the actions performed by participants in the VR experiment to those performed by the victims in a fire can show if they tackle the emergency in the same way. The actions need to be well-defined and easy to identify in both the fire event and the VR experiment (e.g., number of people that tried to extinguish the fire). The VR experiment needs to allow for the participants to have the same choices the people in the fire event had in terms of understanding of the circumstance and objects or tools they can use. The actions performed are considered part of the behavioral dataset.

By focusing on a number of specific actions it is possible to see if the behavior of participants in a VR experiment follows the same logics observed in the fire event. Nevertheless, special care needs to be put on making sure the VR experiment offers the same affordances[2] the physical-world does. Familiarity with the environment, for example, may differ significantly between the fire event and the virtual scenario. Participants in a VR experiment have mere minutes to get acquainted with the virtual environment, a place they probably never have seen before. They may not know the layout; they may not have had enough time to know the way out of the building. People in fire events usually walked into the room where they were when the fire started, and they may have been there for a while if not hours upon hours from frequenting the place every now and then through the years. Therefore, it cannot be expected that the participant in the VR experiment will have the same perception as the fire victim for the sole reason of both environments sharing the same layout.

The example of familiarity shows how differences can arise between the physical and the VR samples due to less prominent aspects of the design of the VR experiment. It is necessary to consider these possible sources of differences in the participants' behavior and adapt the design of the virtual environment in order to compensate or minimize their effect.

3.1.1.2. Walking paths

Walking paths can also be used to compare behavior in VR and in fire events. The walking paths can be tracked in VR continuously, by recording the coordinates of the participant in the virtual environment, in a time-step based approach. Alternatively, an event-based approach can be adopted, in which series of events or milestones are defined. In the event-based approach, the exact location of the participant is not pinpointed at any time. Instead, the sequence in which each event or milestone is reached is recorded.

The comparison of the walking paths also shows how well the behavior in VR reflects the observations made in the physical-world. The analysis of walking paths has its own biases, especially if the locomotion type is not room-scale based. The amount of effort needed for the participant to move around may be

[2] Gibson (1986) defines the affordances of an environment as "what it offers the animal, what it provides or furnishes, either for good or ill. The verb to afford is found in the dictionary, the noun affordance is not. I have made it up. I mean by it something that refers to both the environment and the animal in a way that no existing term does. It implies the complementarity of the animal and the environment."

inversely proportional to their will to do so. Teleportation and even controller-based locomotion require little to no effort to navigate the environment. Motion-based and room-scale-based locomotion require physical effort that participants may be more or less willing to do despite of the situation in the virtual environment. Fatigue or their own estimation of how taxing the movement will be can dissuade them, and those reasons are out of the control of the researcher. Therefore, their reluctance to do the required effort to reach a destination in the virtual environment may affect the result on whether they reach it. Even when using controller-based locomotion, participants that are not fully comfortable using the hand-controllers may default to more simple paths to avoid the struggle.

Another source of bias in the walking paths can be curiosity. Participants are not familiar with the virtual environment, and may be curious and want to explore it. Even if they are familiar with the layout, they may try and look around to confirm that the place in fact is a replica of the physical environment they know.

3.1.1.3. Visual cues

Eye-tracking technology can be added to compare the fixation of the gaze in different objects of interest both in a physical-world environment and a virtual one. This is obviously only valid for physical-world experiments in which participants are equipped with eye-tracking devices.

Comparing the visual cues participants used for finding their way out of the building could help to identify differences in the perception they have of the virtual environment compared to the physical one. Differences between the samples can point out sources of bias in the virtual environment to be corrected.

3.1.2. Selected fire events

The selection process was applied several times until a fire event was deemed suitable for its reproduction in a VR experiment. Some fire events that were considered unfit for a VR experiment are mentioned here with a brief explanation of why, to illustrate why some scenarios were rejected.

- Station Nightclub fire (Grosshandler et al., 2005): conditions like low visibility and crowd pressure could not be replicated appropriately in VR.
- Gothenburg Nightclub fire (Comeau & Duval, 2000): without the consequences of smoke inhalation and risk of injury, jumping out of the window in VR nullifies the problems the fire victims had to face.
- Haunted house fire (Bouchard, 1985): the victims in that fire struggled to realize the flames they saw were not another feature in the haunted house. Therefore, it was assumed that it would be even more difficult for

participants in the VR experiment to do so, without any smell of smoke or the physiological effects of it.

- Beverly Hills Supper Club fire (Swartz, 1979): the waiters played an important role in the evacuation during this fire, instructing the patrons they were serving. Therefore, the waiters in the virtual environment would need a high level of programing and animation for them to interact realistically with the participant before and during the emergency, which was unlikely to be achieved.

Throughout the selection process, the physical-world cases were analyzed from the perspective of behavioral realism. If it were assessed that the VR data could be severely affected by the differences between the physical-world case and the virtual environment, the physical-world case was rejected.

After several iterations, five physical-world cases were identified, that could be replicated in a VR experiment. Table 4 presents an overview of the different events that were deemed feasible based on the three factors discussed on section 3.1. The following subsections will describe each of the experiments in terms of behaviors of interest and features included in the virtual environment to enhance the behavioral realism of the experience.

Table 4 – Description of events deemed feasible with respect to the three factors that are included in this research

Event	Availability of data	Behavior of interest	Replicability in VR
domestic fires	publications from Canter et al. (1980), Sime (1984), and Runefors, Johansson, and Van Hees (2016)	life-preserving actions	3D model of a house, phone to call firefighters, fire extinguisher
MGM Grand fire – guests trapped in the hotel rooms	fire investigation report (Best & Demers, 1982), (Bryan, 1983), NFPA archive of Bryan's questionnaire study	life-preserving actions	interactive smoke layer, working TV, running water
experiments on use of evacuation elevators	raw data from Mossberg, Nilsson, and Andrée (2020) and Andrée et al. (2016)	decision to use elevators, use of signage for decision making, waiting time	working elevator and its lobby, identical to the real building, eye-tracking
hypothetical study on way-finding	self-conducted hypothetical study replicating the conditions of a VR scenario – not yet conducted	compliance with evacuation signage	400 m long tunnel, robot able to find player and give instructions
Stardust nightclub fire	Tribunal report (1982), *Stardust* miniseries (Donnely, 2006)	effect of unresponsive patrons on decision to evacuate	150 non-player characters, fire growth and spread,failed extinguishing attempt

3.2. Defining success

Having presented the behaviors of interest to be measured to then be contrasted to those of the fire event, a definition is needed for what constitutes success. The concepts of failure and success may be seen as complementary, but when assessing how similar two samples are, the limit separating them can be hard to identify. Failure can be defined as missing the target, whichever it may be. The behavioral pattern was not observed; the VR behavioral data had nothing in common with the physical-world one. Success, on the other hand, is nuanced. Was the behavioral pattern observed in *all* participants or only some? Is half of them enough? In exactly how many participants should a specific behavior be observed for it to be considered a typical occurrence in VR? How close should the VR data be to the physical-world data to be considered close enough? Who determines what is close enough? The threshold for success can be debatable.

The uncertainty of what constitutes success when comparing VR data to physical-world data may lead to reach for statistics intuitively, as it is a tool traditionally used to assess whether a control and a treatment are different. Statistical hypothesis testing starts by assuming the null hypothesis is correct, and contrasts the p-value to a significance level to determine whether or not the null-hypothesis needs to be rejected. The significance level sets the threshold of what is considered a low enough likelihood of rejecting the null hypothesis when it was true. The default significance level is typically 0.05 or lower, depending on several factors and sometimes including corrections for multiple comparisons, such as Holm-Bonferroni (Holm, 1979) or any of the many alternatives. However, the significance level only reduces the occurrence of false positives, it does not deal with whether the samples are similar. A possible way of telling how close the two samples are from each other would be focusing on the p-values obtained. A small p-value, independently of the significance level chosen, points at a larger difference between samples. A large p-value, consequently, points at the samples being more similar.

Having identified the p-value as a unit of measure, the next step is to define the threshold for success. Determining how close the VR data needs to be to be considered close enough (that is, how large the p-value should be) is a task for an experts committee in the field. An agreement at community level needs to be reached to assign a standard value to be considered enough. Possibly, a case-by-case analysis is needed instead.

3.3. Procedure

Five VR experiments were conducted in the context of this research work, and they have been numbered from 1 to 5 to identify them. Their findings were published in the four papers included in the research work and in the related publications. The publications corresponding to each experiment give a high degree of detail in terms of virtual environment, scenarios, procedure and analysis of the results. A summary of the experiments performed will be included in the following subsections as background.

All experiments performed in the scope of the present research work consisted of single-player, first-person perspective virtual environments using an HMD and its hand-controllers (HTC Vive series, which was the state of the art at the time). The type of locomotion used, and therefore the equipment used for it, was based on the design of each experiment.

In three out of five experiments participants were recruited from the general public; in the other two experiments specific populations were targeted: residents from a specific neighborhood (see Experiment 1), and employees from the European Organization for Nuclear Research (CERN) or from the European Spallation Source (see Experiment 4). The call for participants from the general public was made through a specialized online platform for researchers from Swedish universities. A large proportion of participants in those experiments were students at Lund University, but no student from the Fire Safety Engineering programs taught at the university were accepted as participants. All VR experiments were conducted in the municipality of Lund, Sweden, except for two scenarios run at CERN in Geneva, Switzerland.

Participants were not told in advance that the VR experiment would be related to any sort of emergency. Upon arrival to the experimental room, they signed an informed consent form. Before launching the scenario, participants were trained on the use of the equipment to ensure they could interact with the virtual environment appropriately. Continuous motion was used in all experiments, but the type of locomotion varied among some of them. Three experiments included non-player characters, either as props or playing an active role in the emergency presented. The participant was not given a body in the virtual environment, but the hand-controllers they were using were identically replicated and tracked in it.

After the VR experience concluded, all participants were presented with questionnaires tailored to the corresponding experiment. Some questions aimed to get participants to explain their reasoning for the way they acted during the experiment. All questionnaires asked to rate some aspects of the VR experience, such as realism, presence, level of stress, fear, discomfort, and difficulty using the equipment. These questions were aimed at evaluating the VR experiment method from the point of view of the experience for the participants. Their levels of stress were self-reported, and no measurements were made of physiological responses to stress, such as heart rate or skin conductance. No additional stressors (such as smell of smoke or radiative heat) were used.

3.3.1. Experiment 1 (Paper I and IV)

The first experiment performed in the context of this research work aimed at studying whether the behavior of participants in a VR domestic fire follows the general model on behavioral sequences developed by Canter et al. (1980). This model is broadly used in the field of Human Behavior in Fire (Fridolf, Nilsson, & Frantzich, 2011; Kuligowski, 2016; Guylène Proulx, 1993), and was therefore considered an acceptable source of data on physical-world behavior. The general model is not the only available model for behavior during fires, but it was selected due to its simplicity.

The model, based on data collected by Sime (1984), covers the expected sequences of behavior of building occupants once a fire is detected in the building. The fires included in the original study took place in three different kinds of buildings: residential buildings, multiple occupancy buildings, and hospitals. Only the data on residential buildings was considered for comparison in the VR experiment. The original dataset was gathered by conducting interviews on victims of residential fires, and their answers were classified and arranged in a decomposition diagram, from which the general model was derived (Canter et al., 1980). The general model was based in different, independent domestic fires in the late 1970s, and therefore it was considered it would apply for any domestic fire. This experiment, thoroughly detailed in *Paper I*, meant to study its application in a virtual residential fire.

Figure 5 - The virtual version of the house replicated in Experiment 1

The virtual environment consisted of the two-story house, its garden and the street where it was located along with other identical houses in a tract housing project. Figure 5 presents a view of the house from the outside. The house was fully furnished, with furniture, appliances and objects commonly found in any home.

The behavior of interest was the sequences of actions performed by the participants, based on the actions codified by Sime (1984). Following the procedure presented in Canter et al. (1980), it would be possible to assess the applicability of the general model on the VR data. Two conditions for success were identified:

a. The behavioral sequences of the participants in Experiment 1 covered all possible sequences described by the general model: the behavioral sequences obtained in VR should be varied, like the original data, and over all paths of the general model.

b. No behavioral sequence is overrepresented: in addition to the first condition, a reasonable distribution of behavioral sequences among the different paths is expected. No single path of the general model can account for most of the behavioral sequences observed.

Two scenarios were produced, identical in any way except for one feature: one had the sound of a domestic smoke alarm as initial cue of a fire, and the other did not have it. Residents of the tract housing project were recruited as participants for both scenarios, being therefore familiar with the house. Additionally, the smoke alarm scenario was run with a sample of participants who did not live in those houses, to observe if familiarity played a role in the participants' behavior.

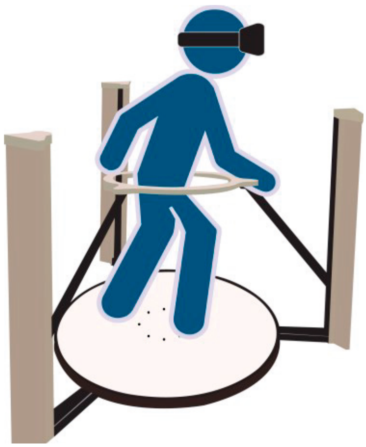

Figure 6 - Illustration of the omnidirectional treadmill used by participants on Experiment 1

The type of locomotion used was motion-based, consisting of an omnidirectional treadmill, as described on *Paper I*. The omnidirectional treadmill did not allow for natural walking, being the motion needed for it to operate somewhat similar to skating. Nevertheless, it allowed 360 degree turns and covered any travel distance, making it possible for the participants to navigate the entirety of the two floors at their preferred speed. The omnidirectional treadmill used in this experiment was selected based on its availability at the time, as access to alternatives was limited. The omnidirectional treadmill is illustrated on Figure 6.

The actions performed by the participants once they became aware of the fire were recorded. The behavioral sequences of all participants were then used to draw a decomposition diagram. More details about the experiment can be found on *Paper I* and *Paper IV*.

3.3.2. Experiment 2 (Paper II and IV)

Experiment 2 was designed to try to recreate the experience of guests in their hotel rooms during the MGM Grand fire as closely as possible, without the life-threatening risks. The virtual environment was based on information gathered from questionnaires used for the investigation on human behavior during the fire of MGM Grand Hotel (Bryan, 1983). The National Fire Protection Association (NFPA) made the questionnaire available upon request for the study presented on *Paper II*. They consisted of a set of questions answered by survivors of the fire, giving detailed information on how they became aware of the fire, where they were, with whom, how they reacted, etc. The questionnaires also gave details about the smoke conditions in their hotel rooms and the hallway, their failed evacuation attempts, how they protected themselves from the smoke, and how long they waited until they could safely leave the building. This trove of first-hand witnesses' accounts was rich and thorough in details about their experiences. It allowed to have a general idea of what the scenario was like for them. With this information, it was possible to create a virtual hotel room as seen on Figure 7 to simulate their experience in a VR experiment, based only on the answers provided by guests who were trapped in their hotel rooms.

One of the questions asked them specifically to indicate the first five actions they did after they became aware of the fire. That detailed account of events was then considered a behavior of interest on actions performed. If participants in the VR experiment performed the same actions in a similar proportion, it could be argued that the VR data matched the physical-world data. In this experiment, the behavior of interest was the actions performed, to be compared to those described by the fire victims in the questionnaires.

Figure 7 - Virtual hotel room used in Experiment 2

Resembling the scenes described by the victims in the questionnaires realizing they were stuck in their room, participants were to figure out what to do about the smoke layer descending on them from the ceiling level. The smoke layer consisted of the physically-based smoke developed by Wahlqvist and van Hees (2018), which was programmed to descend to the floor unless the vent it was coming from was blocked. Independently of how or when the vent was blocked, the smoke layer could be cleared by ventilating the room through open windows.

Special care was put on making the virtual hotel room as realistic as possible, with features such as a working TV with its controller that allowed changing channels, working light switches, faucet and shower. The special attention to these details was not just for the sake of making the room more realistic, but because of the actions that the fire victims performed. Some of them turned on the TV and searched through the channels for information about the fire and what to do. Some of them wetted textiles (like towels, blankets, sheets) to use them as blockage in the gaps around the door or the ventilation ducts, depending on where the smoke entered the room. For the VR experiment to be comparable to the fire, it needed to allow the same actions. Participants, however, were not told in advance about these features in the virtual room. All they were told was that the room was like a real one, which was not true, but it aimed to encourage them to treat the virtual environment as a real one. Those who tried the TV remote controller or the faucet and the shower in the bathroom found out they worked.

Participants performed different numbers of actions, which were recorded. The experiment ended once the participant did all possible actions, or once they cleared the smoke and stood in the room waiting for something else to happen. In some cases, the experiment ended once it became clear the participant did not manage the smoke at all and had no intention to do so.

The locomotion type used in this experiment was room-scale-based. The hotel room was as large as the area of coverage of the sensors of the VR equipment. Participants were equipped with a wireless solution that covered the same area. At some point, the wireless solution stopped working, and the rest of the participants had the HMD connected to the computer through a cable. The cable was also long enough to cover the entirety of the hotel room.

There was a single scenario in this experiment. After the experiment, participants were presented with some of the questions of the original questionnaire (most of those used to create the virtual scenario), to compare them to the answers given by the fire victims. This experiment is detailed in *Paper II* and *Paper IV*.

The data collected on Experiment 2 was compared to the answers in the questionnaires used by Bryan (1983), allowing for a direct comparison between the behavior observed in the VR experiment and that declared by the fire victims.

3.3.3. Experiment 3 (Paper III)

In the third experiment, fully detailed in *Paper III*, the physical-world data originated from a physical experiment about the use of elevators for evacuation in a fire (Mossberg, Nilsson, & Andrée, 2020). The physical experiment was conducted in a high-rise hotel building, and eye-trackers were used to collect data on the visual input participants used to for way-finding. Additionally, Experiment 3 was compared to a study conducted by Andrée et al. (2016), which used a CAVE. All three experiments (the physical experiment, the CAVE experiment, and Experiment 3) were based on the same event using the same layout. Figure 8 shows the hallway of the virtual hotel in Experiment 3.

In Experiment 3, one of the behaviors of interest studied was the decision to use the elevators or the stairs for evacuation, given that the participants were on the 16th floor of the building. Two scenarios were tested, which were identical except in the location of the participant when the alarm was triggered. In the first scenario, the participant was in a room near the elevator lobby. In the second scenario, the participant was in an identical room at the other end of the hallway, near the emergency stairs.

Another behavior of interest was used, based on the eye-tracking data that were collected in the physical experiment. In the physical experiment, the aim was to study how much the participants look at the emergency signage to find their way to the stairs or the elevator in either experiment. In Experiment 3, the same data was collected in order to contrast it with that of the physical experiment.

Figure 8 - Hallway of the virtual high-rise hotel building used in Experiment 3

Lastly, the travel paths were collected using an event-based approach. A series of milestones was identified and the sequence in which the participant reached each milestone was recorded. A similar analysis was made with the travel paths of the participants in the physical experiment, making the two samples comparable.

The CAVE experiment (Andrée et al., 2016) collected data on the waiting time for the elevators. In the CAVE experiment, participants were left waiting for the elevators to come to their floor for up to 20 minutes, although around 60% of them waited for five minutes or less. This waiting time was not possible to be collected in the physical experiment, as the physical experiment took place in a single floor while the rest of the occupants in the high-rise building were unaware of it. Therefore, the elevators were operating as usual, and the physical experiment ended the moment the participant pressed the elevator call-button to avoid them alarming other building occupants about an emergency. In the CAVE experiment, and subsequently in Experiment 3, the elevators moved up and down constantly, never stopping at the floor where the participant was. In Experiment 3, participants were left waiting for up to five minutes.

The waiting time cannot quite be considered a behavior of interest in the context of this work, as the comparable data was obtained in a different VR experiment, not a fire event. Nevertheless, this data was compared with each other in order to contrast the equipment used and the effect it may have in the results.

Controller-based locomotion was used in this experiment. Participants had to press a button in the hand-controller to command movement. The walking speed was constant at around 1 m/s. The direction of movement was controlled by the orientation of the HMD: the participant had to turn their head to the desired direction of movement. As the locomotion was controller-based, the standing-only configuration of the VR equipment was used, meaning that the participant did not get to move in the physical room. They were encouraged to turn their bodies as they saw fit, to avoid uncomfortable neck positions when changing the direction of movement. *Paper III* offers more details about Experiment 3.

3.3.4. Experiment 4 (Paper IV)

A VR experiment was conducted on way-finding systems for the evacuation of a particle accelerator, is described in a related publication (Arias et al., 2019) and brought up in *Paper IV*. A virtual environment based on the Large Hadron Collider (LHC) was produced for a study in collaboration with the European Organization for Nuclear Research (CERN). The virtual particle accelerator shown on Figure 9 was not a replica of the LHC but it was based on it. It had the dimensions of a projected, much larger collider, as this experiment was a minor part in a major international collaboration: the Future Circular Collider study (Abada et al., 2019). The VR experiment aimed at testing different way-finding systems to optimize safe evacuation in a long, underground accelerator tunnel. There were three scenarios with different levels of sophistication in their way-finding systems, two of them including a robot specialized for evacuation purposes. The robot was based on an existing train inspection monorail robot in the LHC which moves on a rail at ceiling level. A virtual version of this robot was adapted for evacuation purposes with the support of the robotics team at CERN, which controls and maintains the existing robot at the LHC. The existing robot, however, does not have evacuation tasks at the moment, although this functionality could be added eventually for the Future Circular Collider, which is the reason why the robot as considered an option.

Figure 9 - View of the particle accelerator tunnel used in Experiment 4

The robot in the virtual environment was designed to move on its rail in the same direction as the participant, but to instruct them to evacuate in the opposite direction. It was considered that the evacuation system had to be robust enough to convince the participant to change their direction of movement, even if that meant to walk back to the starting point. Employees at particle accelerators were recruited as participants. The chosen locomotion type was controller-based locomotion, due to the length of the segment of the virtual accelerator tunnel (around 400 m long). More details about the scenarios and the results can be found in the associated publication (Arias et al., 2019).

This experiment was to be contrasted with data collected from a hypothetical study in the form of an online survey. The participants of the online survey were to be employees at CERN. The online survey was designed to show short videos made in the same virtual environment used in the VR experiment. The participants were meant to watch a video (around 20 s) after which a decision was to be made. In the first video, they were to see the moment the alarm went off, and be asked to make a decision of which direction to go. A second video would show them the robot and the message it gave, and ask them then whether they would continue going in the same direction or turn back. Up to four decisions were to be made by the participant, with the hypothetical study ending the moment they chose to turn back. However, due to time constraints, the corresponding comparison between the two sets of data could not be included in this research work. Nevertheless, important lessons were learned about limitations of VR when applied for evacuation experiments, and therefore Experiment 4 constitutes an important source of knowledge in this research work in the fulfillment of the third objective. Experiment 4 is discussed on *Paper IV*.

3.3.5. Experiment 5 (Paper IV)

The fifth experiment was based on the Stardust Nightclub fire, which took place in Dublin in 1981, resulting in 48 deaths and over 200 injured (Tribunal of inquiry of the fire at the Stardust & Keane, 1982). The VR experiment, as described in a related publication (Arias et al., 2018), was no exhaustive replication of the events during that fire. The general layout of the available rooms in the nightclub that evening and the location of the fire were replicated. Some other details were added, such as the timing for the fire growth and spread, following parts of the timeline presented by the Tribunal report. Although smoke was included in the virtual environment, the smoke spread and visibility conditions of the fire were not replicated since they would only make the virtual environment too dark. As explained before, the VR experiments run in the context of this research work relied heavily on visual input, and low visibility conditions would restrict severely the participants' perception of the surroundings. Therefore, it was decided to keep good visibility conditions by generating a smoke layer that did not reach the participant at any point. Figure 10 depicts the virtual environment used in this experiment.

Figure 10 - View of the virtual nightclub used in Experiment 5. The image has been edited to look brighter. The original virtual environment was darker, as a real nightclub

The behavior of interest was the decision to evacuate. It is known that in the fire, many people did not evacuate immediately after realizing there was a fire (Tribunal of inquiry of the fire at the Stardust & Keane, 1982). Several factors may have played a role (such as: the music that kept playing for a few minutes after the fire was discovered, people were watching the extinguishing attempts, possibly underestimating the fire growth rate, etc.). Experiment 5 aimed at identifying at which point a participant in it would decide to evacuate. Four different key milestones were identified, that would indicate the moment they decided to evacuate: when the participant realized there was a fire, when they saw the failed attempts to extinguish it, when other patrons left, or only after the disk jockey stopped the music and gave a voice message. These milestones followed the description of the events presented in the investigation report (Tribunal of inquiry of the fire at the Stardust & Keane, 1982).

The virtual nightclub included over 150 non-player characters dispersed it, seated at the tables, standing in small groups simulating to talk to each other, and dancing. The experiment consisted of two scenarios that were identical except for the information participants received before starting. Half of the participants were told that an additional participant may join the virtual environment from a remote laboratory in another country. It was explained to them that scheduling conflicts made it impossible to tell if or when the remote participant would join. That was a purposely vague information, as there was no remote laboratory or additional participant whatsoever. The scenario was still a single-player one, but the participant was made believe it could be multiplayer. The false multiplayer scenario was an attempt to observe if participants would behave differently if they thought another human may be around in the virtual environment.

Controller-based locomotion was implemented in this experiment. The decision to evacuate was recorded, based on the four milestones described before. It was only considered that the participant evacuated once they reached for the handle of one of the emergency exits. Once the experiment concluded, participants in the false multiplayer scenario were asked if they saw the remote participant, and their answer was noted. During the debriefing session, they were told there was no remote participant, and were asked if they believed the researcher when they claimed they might be one, in order to assess if the deception worked. This experiment is detailed in Arias et al. (2018) and discussed on *Paper IV*.

3.4. Summary of the VR experiments performed

The VR experiments performed in the scope of this research work varied in many ways: type of fire event, type of building, number of scenarios, and number of participants, among others. Table 5 indicates the number of scenarios and total number of participants as well as the publications that reported the results in the scope of this research work.

Table 5 – Summary of the VR experiments run in the context of this research work and the physical-world cases they replicated. *The hypothetical study was not finished at the time of the writing of this research work

Experiment	Physical-world case	Virtual environment	Scenarios	Participants	Publication
1	various residential fire	two-storey house	3	66	Paper I & IV
2	MGM Grand fire	hotel room	1	55	Paper II & IV
3	use of evacuation elevators	high-rise building	2	62	Paper III
4	hypothetical study*	particle accelerator	3	110	Paper IV
5	Stardust Nightclub fire	nightclub	2	67	Paper IV

4. Discussion of results

The results discussed in this chapter are classified in relation to their role in the fulfillment of each of the objectives presented in Chapter 1. Additionally, results on the assessments made by participants on the VR experience will be presented.

A considerable part of these results has been presented in the corresponding publications included in the research. However, the discussion in the following subsections focuses on what those results mean from the perspective of the VR experiment method.

4.1. Behavioral patterns

The first objective of this research, objective 1, aimed to "investigate if the behavior of participants in VR experiments follows the same patterns reported in fire incidents". To fulfill this objective, the behavioral patterns obtained from the different VR experiments need to be similar to those reported in fire incidents. Those physical-world patterns refer to the different theories used to explain evacuation behavior in Human Behavior in Fire, as presented in Chapter 3. The following subsections describe the results observed on behavioral patterns in the different VR experiments, followed by an in-depth discussion of their role in the fulfillment of Objective 1.

4.1.1. Results on behavioral patterns

The results on behavioral patterns presented here are described in the context of the corresponding VR experiment the behavioral pattern was more notorious. Experiment 1 was specifically designed to observe the occurrence of the behavioral sequences described by Canter et al. (1980). None of the other VR experiments was designed with the objective of observing a specific theory or model of behavior.

4.1.1.1. Behavioral sequences

In Experiment 1, the VR data was used to generate decomposition diagrams similar to those produced by Canter et al. (1980). Figure 11 reproduces the general model (Canter et al., 1980). Figure 12 and Figure 13 present the decomposition diagrams generated using the VR data, as presented in *Paper I*. As shown, the VR dataset did not only reproduce all possible sequences of actions from the general model, but more importantly there was no specific sequence was overrepresented. The similarities cannot be quantified in this case, but they are clear from the perspective of the behavioral patterns. Participants in the VR experiment showed the same type of behavior that would be expected from victims in a fire. There was misinterpretation of the alarm (in the scenario that had an alarm), as well as exploration, attempts to fight the fire, warn others, and finally evacuate. No branch of the general model was absent in the decomposition diagrams produced.

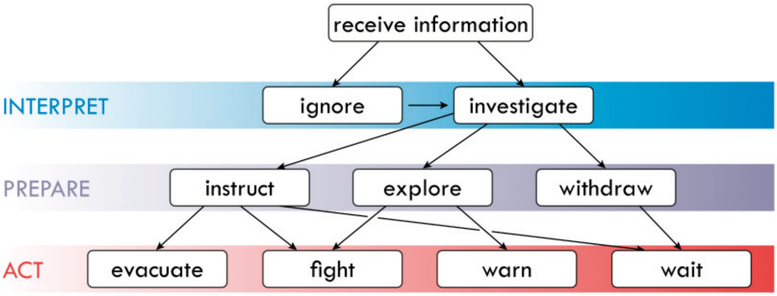

Figure 11 – Reproduction of the general model by Canter et al. (1980)

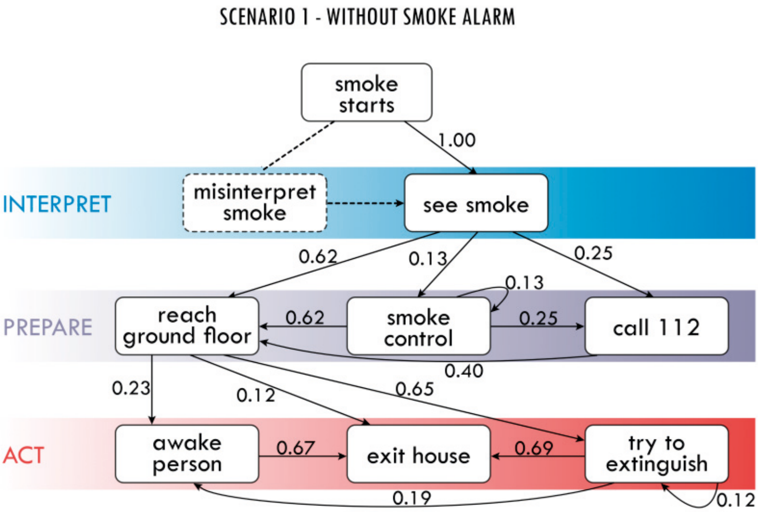

Figure 12 – Decomposition diagrams obtained from the VR data from scenario 1 in Experiment 1, reproduced from Paper I. The numbers on the arrows represent the proportion of participants who performed a given action moving on to perform the following one

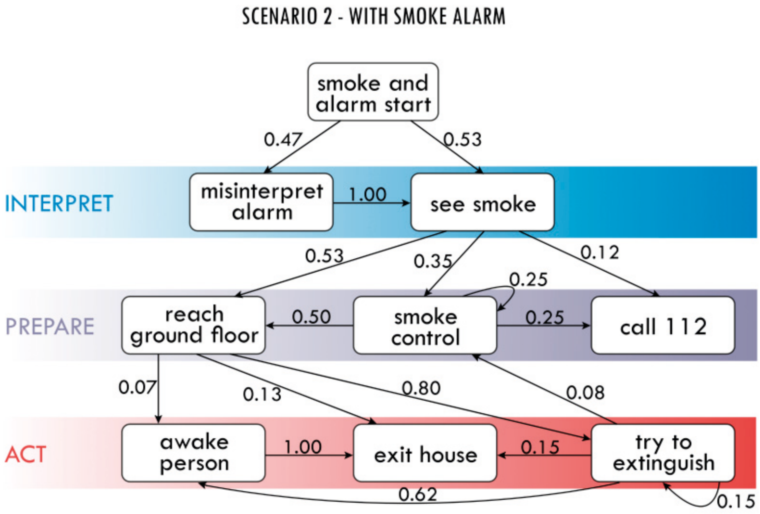

Figure 13 – Decomposition diagrams obtained from the VR data from scenario 2 in Experiment 1, reproduced from Paper I. The numbers on the arrows represent the proportion of participants who performed a given action moving on to perform the following one

4.1.1.2. Repeated attempts

As the behavioral diagrams produced by Canter et al. (1980) indicate, in some cases people in fire emergencies make repeated attempts at the same action. This was also observed in participants in Experiment 1, as shown in Figure 12 and Figure 13. The sequence of actions is indicated by arrows with labels. An arrow starting in action A and ending in action B means that participants performed action A and then action B. The label on each arrow indicates the proportion of participants who performed action A, and action B subsequently. In some cases, participants performed one action, and a returning arrow shows they performed it again. In these cases, participants did not simply repeat the action, but attempted a different way to perform it. This can be exemplified by the repeated attempts at extinguishing the fire, described as follows.

Participants in Experiment 1 showed enough engagement in the VR experience to attempt to tackle the fire using other means once the extinguisher ran out. In the design of the virtual environment it was expected that participants would use the extinguisher on the flames, and naïvely, no other means were expected to be used. However, participants showed inventiveness. Some attempted to grab a carpet from the hallway or a towel from the bathroom to put out the fire. However, the virtual environment did not allow for that. Some objects in the house (such as curtains, lamps, furnishings, kitchen appliances) were there only for aesthetical reasons, and the participant could not pick them up or interact with them in any way. This was done for simplicity in the design phase. Therefore, their attempts to use fabrics to extinguish the fire failed. Other participants took a decorative bowl placed on the kitchen table and tried to fill it up with water from the tap in the kitchen, but no water came out of the faucet. Both alternative attempts at extinguishing the fire, even if unfruitful, were registered as a repeated attempt at fire suppression. This VR experiment showed that if objects are available, participants may try to use them. This is especially valid for objects that are commonly used in the physical world to extinguish a fire in absence of a better option. Using water or a carpet to try to extinguish the fire is not unreasonable in fires, and therefore a realistic virtual environment should allow those actions.

4.1.1.3. Unprompted complex actions in VR

In Experiment 2, the participants' behavior showed that complex actions can also be observed in VR. Participants received no instructions about what was possible to do in the VR hotel room. Instead, they learned themselves while exploring the room and by trying different objects before the smoke was triggered. Even without receiving any instructions, 45% of the participants figured out they could block the vent by covering it with some of the furnishings of the room. Moreover, wetting those objects first is here considered a complex action, as it required the participants to remember there was running water in the bathroom and to realize that water may be combined with other objects. This complex action does not require much thought in reality, as people know water will wet an object upon contact. However, as explained on section 4.1.1.2, it is possible that in a virtual environment some objects are not designed to be used. Nevertheless. some participants tried to wet some materials in the virtual environment and succeeded at it. Whit this action, unknowingly mimicked what the fire victims did.

4.1.1.4. Commitment and role-rule model

The effect of commitment in an activity and the Role-rule model were observed in Experiment 4. In that experiment, participants in a virtual particle accelerator tunnel were told to put a given set to tools into a toolbox (shown on Figure 14) before exploring the virtual environment. As mentioned before, that instruction was given at the beginning of their VR experience to distract them, and it was presented as a training opportunity for the use of the hand-controllers. The task was simple and relatively quick to do. The goal was to switch their attention from the experimental room to the virtual environment. The placing of the fifth tool into the toolbox triggered the alarm, catching the participants by surprise. By then, they still had a handful of tools on the table, which seems to have created a conflict in some of them. Most participants did not start their evacuation until the last tool was in the toolbox, with some of them in a clear rush to finish once the alarm went off. Commitment and the role-rule model may explain why.

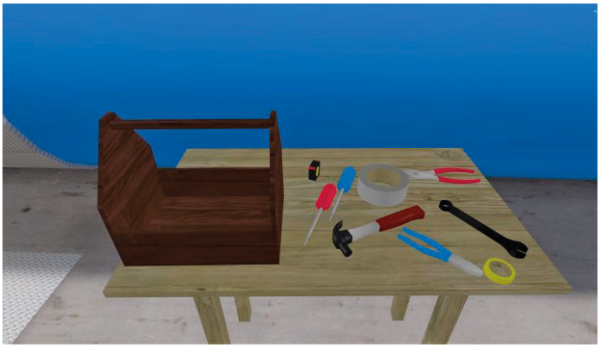

Figure 14 - Tools and toolbox used in Experiment 4 for the bogus task. Placing a fifth tool into the toolbox triggered the alarm

Commitment has been identified as a factor influencing occupant behavior in an evacuation (G. Proulx, 2001). According to (G. Proulx, 2001), “Occupants who have paid good money to watch a trendy movie are not prepared to leave while they are engrossed in the story. They are committed to the activity of watching this movie and the fire alarm signal by itself is unlikely to be sufficient to make them leave”. The task was brief and simple, and by the time the alarm went off, participants were halfway through. Being committed to finish the task is a possible explanation for why they would not leave until done. They had the intention of finishing the given task before the alarm, and the alarm on itself did not seem to be enough to change that decision.

It is possible that, more than just commitment, the fact that the task was given to them by the researcher was received as an instruction to comply with. The role-rule model (Tong & Canter, 1985) claims that people’s roles before an emergency starts continue to play once it started. As presented by Tong and Canter (1985), “this model postulates that people’s conduct is guided by a set of expectations they have about their purpose in a particular context. The general framework formed by these expectations is known as their ‘role’. The activities they engage in to fulfil their role are influenced by guiding principles or ‘rules’”. In the context of the experiment, a person who volunteers to participate takes the role of the participant. As such, the rule for that role is to follow the instructions given by the researcher. In the same context, the role of the researcher is to instruct the participant on what they are supposed to do during the experiment. In the case of the task given in Experiment 4, the instruction about putting all tools into the toolbox clashed with the instruction to evacuate implied by the alarm. By design, the alarm being triggered in the middle of the task intended to make the emergency start unexpectedly. It was irrelevant to the study whether they put all the tools into the toolbox, but participants were not aware of that. Therefore, finishing the task first and in some cases much quicker than before the alarm started, was the fulfillment of their role. This behavior was not expected, but it indicated that the bogus task given can create a contradiction between the instruction they received and the alarm signaling they had to evacuate.

In the case of Experiment 4 the task was relatively simple, and even though many participants did not start their evacuation until the last tool was into the box, finishing it took them only a few more seconds. Moreover, it is not uncommon to evacuees' behavior to finish what they were doing before leaving, and commitment has been seen in evacuations even when the emergency is obvious (Frantzich, 2001). The pre-evacuation time was not measured in this experiment, and did not play a role in the data collected. However, it should be considered that a bogus task that takes too long to complete, may in fact affect the participants' response to the emergency. Since the experiment was not designed specifically to measure the occurrence of this behavior, it is unclear whether the alarm being triggered at the end of the task would have shown different results in their pre-evacuation behavior.

4.1.1.5. Theory of affiliation (5)

The theory of affiliation (Sime, 1985) can explain some of the behaviors observed in Experiment 5. In this experiment, the participant was in a nightclub crowded by non-player characters. Half of the participants were told about the possibility of an additional participant joining in the experiment from a remote location (i.e., the VR experiment being multiplayer). As mentioned before, there was no such remote participant. The ruse was introduced in an attempt to counteract the anti-social behavior seen in Experiment 1. The behavior of participants in the supposed multiplayer scenario could show whether participants take non-player characters more seriously if they thought one of those non-player characters could be a real person.

Many participants in the supposed multiplayer scenario invested time in searching for that other human in the virtual environment, once they became aware of the fire. This behavior can be seen as an application of the Theory of Affiliation (Sime, 1985). According to that theory, people in evacuation situations tend to move towards the familiar: familiar places, familiar people. In Experiment 5, the participant may not have been familiar with the supposed remote participant, but what made them similar is their human nature, contrasted with a crowd of non-player characters that were unresponsive to the participant.

4.1.2. Fulfillment of objective 1

As presented in Chapter 1, objective 1 in this study is to "investigate if the behavior of participants in VR experiments follows the same patterns reported in fire incidents”. Results presented on section 4.1.1 showed that certain behavioral patterns emerged in the VR experiments. As it is expected from fire victims, participants in different experiments attempted different ways to tackle the emergency they were facing, and were able to perform complex actions in the virtual environment.

Some behavioral theories and models used to explain the behavior of building occupants during a fire were able to explain behaviors observed in the VR experiments presented here. The theories and models used, commitment (Frantzich, 2001; G. Proulx, 2001), the Role-rule Model (Tong & Canter, 1985), and the Theory of Affiliation (Sime, 1985), were used to explain the participants’ behavior in a similar way as they are used to explain fire victims behavior. The fact that these theories and models were still useful to describe the behavior of participants in the virtual environments indicate the fulfillment of objective 1. It should be noticed, though, that the experiments were not designed to test if the theories would apply or not. Therefore, if some behaviors predicted by certain theories did not emerge, at this stage it can only be considered as absence of evidence. Nevertheless, the data obtained indicate that behavioral theories in human behavior in fire also apply in VR experiments, even when the experiments are not specifically designed to prompt that specific behavior.

This research is clearly of exploratory nature. Therefore, it is important to reflect on the falsifiability of its results. As discussed on *Paper I*, variation in the sequence of behavior between participants is an indicator of success. If all participants or the vast majority of them in a given experiment behaved in an identical manner, it would be clear that the virtual environment did not allow them to have any variability in their actions. Not all participants in these VR experiments behaved in the same way, as it can be attested by the results presented on section 4.1.1. The behavioral patterns observed were not present across all five VR experiments, in the same way as some fire victims only show some kinds of behavior due to their background, ability and subjectivity.

The results obtained on behavioral patterns cannot prove that behavioral patterns in VR experiments are similar to those in fire events. However, they did not disprove a possible similarity. A uniform response from all participants across the different VR experiments would be clear indicator of unusual behavior compared to physical-world observations. That would mean 100% of the participants showing the exact same behavior. However, that uniform distribution of response is a very extreme case. It is not clear what would be a credible proportion of participants showing the same response. If 100% of them behaving the same way indicates that the VR experiment did not allow for subjective differences between participants, what would an 87% of uniform responses mean? Is 87% high enough to be considered extreme? What would be the assessment for a 53% of identical responses? Without a clear threshold for when a distribution of behaviors becomes extreme for a physical-world scenario, it is not possible to precisely tell how credible the responses were in the VR experiments. These five VR experiments may not have had enough sensitivity to detect less extreme distributions of responses (anything less than 100% uniformity) that would disprove the similarity in behavioral patterns in VR experiments and in fire events.

4.2. Behavioral data

Objective 2 aimed to "compare behavioral data obtained in VR experiments to that obtained using other research methods". To do so, the behavioral data collected in the VR experiments is compared to the corresponding physical-world data to assess their similarities. As a reminder, the term *physical-world* in the context of this research means *non-VR*. This implies that data obtained from physical experiments is physical-world data. The following subsections describe the results on behavioral data obtained from the VR experiments, and present a discussion on their role in the fulfillment of Objective 2.

4.2.1. Results on behavioral data

The behavioral data collected is here presented in terms of actions performed, walking paths, and eye-tracking data. The sets of VR data are compared with the corresponding sets of physical-world data obtained from other research methods.

4.2.1.1. Actions performed

The relevant actions performed by participants in Experiment 1, Experiment 2 and Experiment 3 were compared to the corresponding behavior in the physical-world datasets. The other two experiments did not have measurable physical-world data to be compared quantitatively at the moment of writing of this work.

The number of actions performed in Experiment 1 is presented on Figure 15. These actions are the six in the *prepare* and the *act* stages of the decomposition diagrams obtained in the experiment (see Figure 12 and Figure 13). Since some participants performed some actions more than once, the total number of actions is larger than six for them. Roughly 80% of the participants performed three to six actions, with an average of 4.15 actions per participant. Eight participants (20%) interrupted the experiment voluntarily due to experiencing nausea or other motion-sickness symptoms. From them, four performed no action, and the rest between three and five actions each. The number of actions performed per participant cannot be compared to the actions reported by the victims in the original dataset (Sime, 1984), as it does not provide this level of detail.

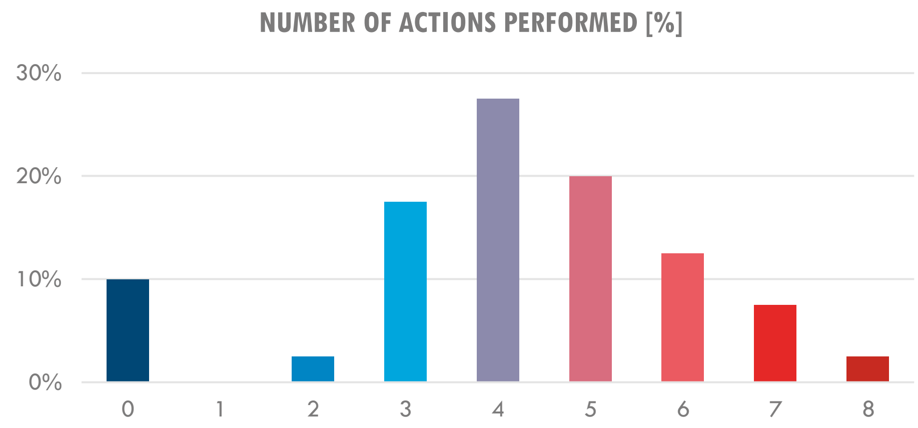

Figure 15 - Proportion of participants performing a given number of actions in Experiment I

In Experiment 2, a direct comparison can be made between the VR and the physical-world datasets. The proportion of participants performing each of the relevant actions was contrasted with the data collected from the questionnaires in the original investigation (Bryan, 1983). Five actions were considered relevant as they were explicitly mentioned in the questionnaire. The proportion of participants and victims performing the different actions is shown on Figure 16. Three out of five actions showed no statistically significant difference between the two samples, as described on *Paper 2*. In the other two actions (block vent and turn on TV), the difference between the two proportions is very large.

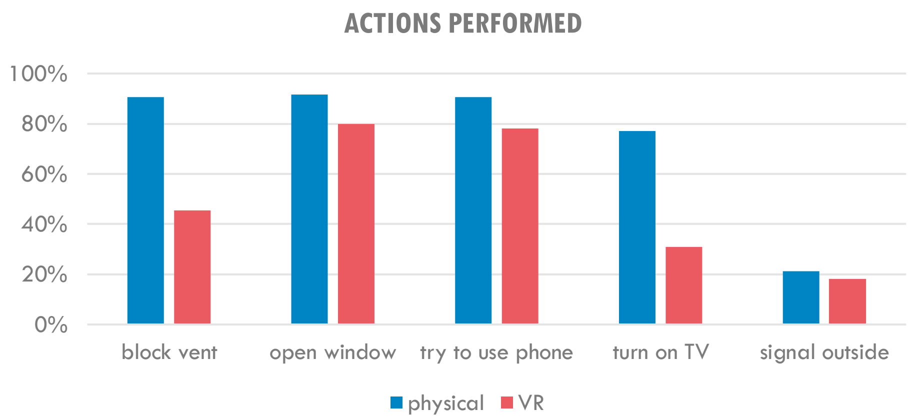

Figure 16 - Proportion of participants in Experiment 2 and physical-world victims performing a relevant action

The number of actions performed by participants in Experiment 2 can be compared to the number of actions performed by the fire victims, as shown on Figure 17. The average number of actions performed by the VR sample is 2.56, and it is 3.05 for the physical sample. The difference between the two averages rather small. Therefore, the behavior of the two samples was not very different.

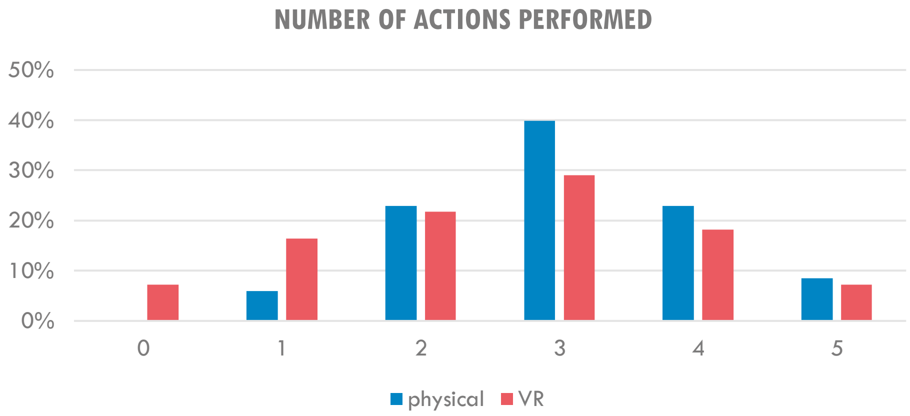

Figure 17 - Proportion of participants in Experiment 2 and victims of the physical fire performing a given number of actions

Table 6 – Number of participants performing each number of actions for both the VR sample and the physical fire victims, and the p-value obtained from the statistical test performed

Number of actions	Physical		VR		p-value
	yes	no	yes	no	
0	0	118	4	51	0.0095
1	7	111	9	46	0.0451
2	27	91	12	43	1
3	47	71	16	39	0.1803
4	27	91	10	45	0.554
5	10	108	4	51	1

In Experiment 3, data was collected on the participants' preference for usage of the two available means of egress: an evacuation elevator and an emergency staircase. This data was compared to the data collected in the corresponding physical experiment, and it is shown on Figure 18. A Fisher's exact test was performed to compare the samples, resulting in a p-value of 0.3001, which means there is a level of similarity between them.

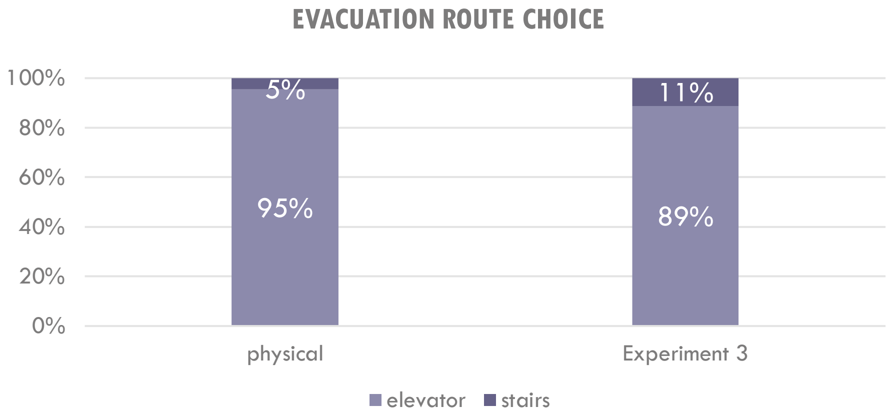

Figure 18 - Preference for the two available means of egress for the three experiments compared in the context of Experiment 3

Participants in Experiment 3 were asked afterwards what they would have done differently had the emergency occurred in reality instead of VR. Over half of the participants claimed they would not use the elevators in reality, which constitutes a discrepancy from the results of the physical experiment, where 95% of the participants went straight for the elevators.

4.2.1.2. Walking paths

The event-based approach explained in section 3.1.1.2 was chosen to collect walking paths in the virtual high-rise hotel used on Experiment 3. The data was collected through visual analysis of the video recordings of each participant. Then, participants were grouped based on their sequence of milestones, putting identical ones in the same group.

There was no identical sequence observed in a large proportion of participants. As explained before, there were two scenarios in this experiment, which differed only in the room participants were assigned to go to: participants in Scenario 1 were assigned to room 1, and participants in Scenario 2 were assigned to room 2 (see Figure 19 and Figure 20). Participants starting in room 2 were more prone than those starting in room 1 to walk down the hallway, past the elevator lobby, and check the dead-end. Interestingly, some participants who started in room 2 in the physical experiment followed the same pattern. It is possible that participants were interested in exploring the virtual environment before making a decision.

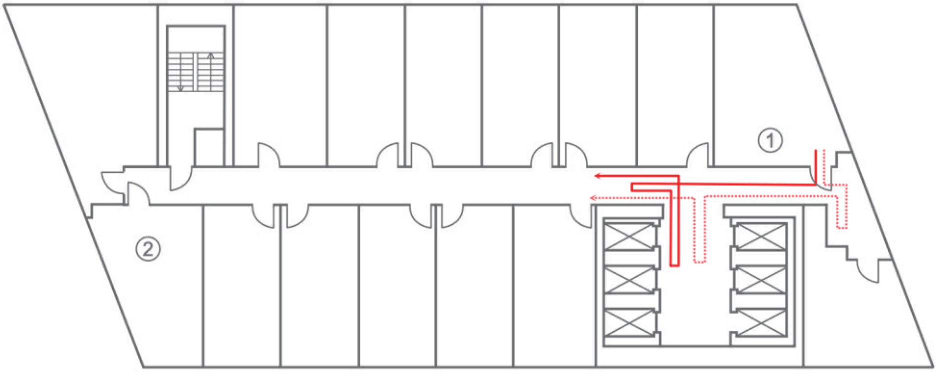

Figure 19 – Typical walking paths for participants in Scenario 1 in Experiment 3. Reproduced from Paper III.

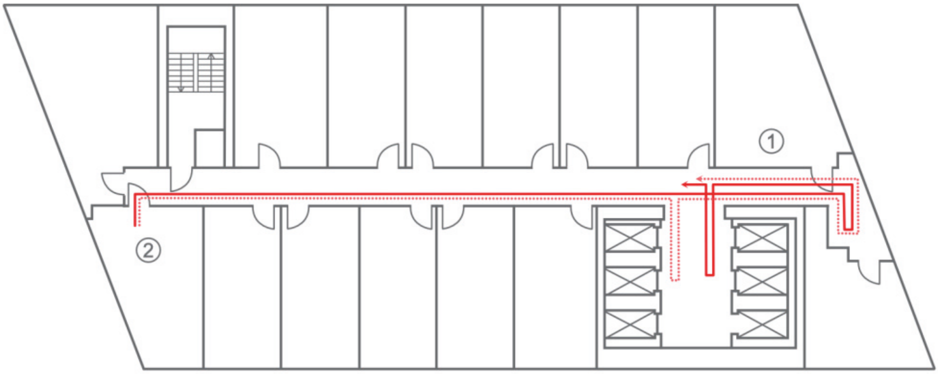

Figure 20 – Typical walking paths for participants in Scenario 2 in Experiment 3. Reproduced from Paper III

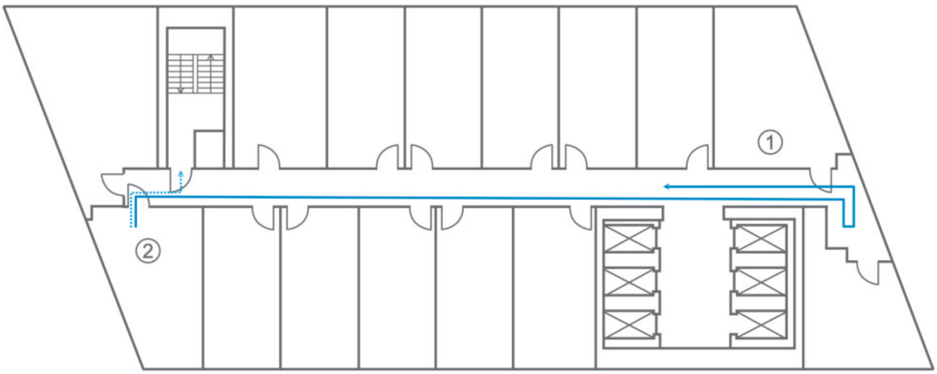

Figure 21 – Walking paths observed in the physical experiment in participants who chose the stairs. Reproduced from Paper III.

The physical experiment also collected data on the walking paths through visual analysis of videos and a set of milestones. Walking past the elevator lobby was observed in participants in the physical experiment. A qualitative assessment of the walking paths from the VR and the physical-world samples showed no clear difference between them. Figure 21 presents the walking paths of the two participants in the physical experiment who chose the stairs for evacuation. One of them went straight to the stairs (as participants in Experiment 3 did as well), and one did some exploration of the hallway first.

4.2.1.3. Eye-tracking data

Three objects of interest were selected in Experiment 3. The eye-tracking device was able to indicate what the participant was looking at, at any moment in the VR experience. However, only glances at specific objects (the objects of interest) were logged, as they were relevant to the purpose of the experiment. The frequency and duration of glances at each object of interest were recorded.

The objects of interest were the emergency exit signs in the hallway on floor 16. These signs pointed at the available evacuation routes: one sign pointed at the elevator lobby from the hallway, another one was on the elevator lobby door (as they closed shut once the alarm was triggered), and a third one pointed at the stairs. Table 7 presents the proportion of participants in Experiment 3 and the physical experiment that looked at each object of interest. According to the data collected, participants in Experiment 3 consistently looked more at the signage than those in the physical experiment.

The p-values obtained showed mixed results in the comparison between the two samples. Part of the variability could be due to the equipment used. Experiment 3 and the physical experiment used different brands of eye-tracking devices. Additionally, differences between the virtual environment and the physical one in terms of brightness, lighting and type of signage, may also explain the differences, as it is discussed in the following section.

Table 7 - Proportion of participants in Experiment 3 and the physical experiment looking at the emergency exit signage

			looks at sign, n (%)		
Object of interest	*Scenario*	*Experiment*	**yes**	**no**	*p-value*
sign pointing at elevator lobby	1	physical	7 (37)	12 (63)	0.7685
		Experiment 3	14 (45)	17 (55)	
	2	physical	9 (41)	13 (59)	0.0226
		Experiment 3	23 (74)	8 (26)	
sign on elevator lobby door	1	physical	13 (68)	6 (32)	0.4963
		Experiment 3	25 (81)	6 (19)	
	2	physical	11 (50)	11 (50)	0.0142
		Experiment 3	26 (84)	5 (16)	
sign pointing at stairs	1	physical	4 (21)	15 (79)	0.3513
		Experiment 3	11 (35)	20 (75)	
	2	physical	8 (36)	14 (64)	1
		Experiment 3	11 (35)	20 (65)	

4.2.2. Fulfillment of objective 2

The second objective of this research work is to "compare behavioral data obtained in VR experiments to that obtained using other research methods". The behavioral VR data was compared quantitatively to physical-world data, and the comparison gave a wide range of results. In some cases there was a large difference between the VR and the physical-word samples. In other cases, there was no difference between the samples. With such a wide spectrum of results, it is difficult to give an overall assessment on whether or not there was an agreement between the VR and the physical-world datasets. However, upon close inspection, the different results individually show strengths and weaknesses of the VR experiment method, as presented as follows.

The average number of actions performed in Experiment 2 by the VR sample and by the physical-world sample are similar. Furthermore, the distribution of participants performing the different number of actions follows, in a reasonable way, the distribution of the physical-world sample, as seen on Figure 17. These similarities show that the behavioral data obtained in the VR experiment did not diverge much from the physical-world data. There is no physical-world data to compare the distribution of number of actions performed by participants in Experiment 1 (see Figure 15). Nevertheless, it is remarkable that the distribution is not too different than the one obtained in Experiment 2 (see Figure 17). Naturally, this comparison cannot be considered very stringent, given the many differences between the VR experiences offered by each experiment.

The specific actions performed in Experiment 2, as shown in Figure 16 and Table 6, generated generally low p-values (low level of similarity between samples). The large differences among the samples in "block vent" and "turn on TV" can be explained by the lack of high stakes in the VR experiment, detailed in section 4.3.1. Moreover, the one action that showed a relatively high p-value (the "signal outside" action, with a p-value of 0.6902) seems to be a result of chance, as explained on *Paper II*. In short, participants in the VR experiment and the victims in the fire event had different motivations to signal their presence or not. On one hand, in the virtual environment, there was nobody on the streets for participants to communicate with. On the other hand, victims of the fire did not need to signal their presence in their hotel rooms (even though some did), since they were clearly visible standing next to the broken windows in broad daylight, watching the rescue efforts. Therefore, the similarities between the two samples are most likely driven by chance rather than similar motivations.

Even though the other two actions recorded on Experiment 2 (i.e., "open windows" and "try to use phone") showed low p-values (less than 0.05), the proportion of participants in each sample performing them differed greatly. The low p-values can be explained by the relatively small sample size. As an example, in the case of the "open windows" action, six more participants performing it would have given a p-value of 1. This means that a change of less than 10% of the responses would have swayed the result considerably. The problem with the sample sizes was a recurrent one given the constraints in time and resources to increase participant numbers. This problem is not unique to VR experiments, as it affects other research methods that rely on voluntary participation from the public (e.g., laboratory experiments and hypothetical scenario experiments). The ubiquity of the problem is no excuse for its occurrence, but it is worth to highlight that it is not exclusive to the VR experiment method.

The comparison of the means of egress used by participants in both Experiment 3 and the physical experiment showed similarities between the samples. Interestingly, the self-prediction made by the participants in Experiment 3, with 50% of them claiming they would not have used the elevators as first option if the emergency would have occurred in reality, was not sustained by the actions of participants in the physical experiment. This result could be a starting point for looking into how VR experiments could be better than hypothetical scenario experiments to collect behavioral data. Asking a single hypothetical question can hardly constitute a hypothetical scenario experiment, but the difference between participants' answer to that question in Experiment 3 and the behavior observed in the physical experiment is too large to be glossed over. As indicated by Kinateder, Ronchi, Nilsson, et al. (2014), there is a lower level of ecological validity in hypothetical scenario experiments compared to VR experiments and classical laboratory experiments (see Table 3). The large discrepancy between the data from the physical experiment and the hypothetical question, and the agreement between the VR sample and the physical sample, support the assessment made by Kinateder, Ronchi, Nilsson, et al. (2014).

Lastly, the eye-tracking data showed that participants in Experiment 3 looked at the emergency signage more often than participants in the physical experiment. This overrepresentation of the use of emergency signage can be due to two factors: the reduced spatial of orientation in VR, and the overall visibility of the signage. As mentioned before, VR users find it harder to orient themselves in a virtual environment than in a physical one (Nguyen-Vo, Riecke, & Stuerzlinger, 2017). Participants in Experiment 3 may have been searching for emergency signage to find their way out, while participants in the physical experiment did not need the signage as much. This means that the VR technology may not provide enough information for some cognitive processes to be completed successfully. Therefore, participants in a way-finding experiment in VR may need adequate aid to compensate for the expected reduction in spatial orientation.

The second reason that could explain the overrepresentation of people looking at emergency signage is the visibility of the signage. Figure 22 shows the virtual hallway participants were in in Experiment 3 and the corresponding hallway in the existing building used in the physical experiment (Mossberg, Nilsson, & Andrée, 2020). There are clear differences between the two images, especially with regards to the size of the signage and the lighting in the hallway. At the time of the generation of the virtual environment, these differences were not deemed important. However, given the differences between the VR and physical-world eye-tracking data, the effect of the visibility conditions cannot be ignored. It is clear, though, that more rigor is needed when replicating a physical environment in VR to truly maintain a one-to-one similarity.

a) b)

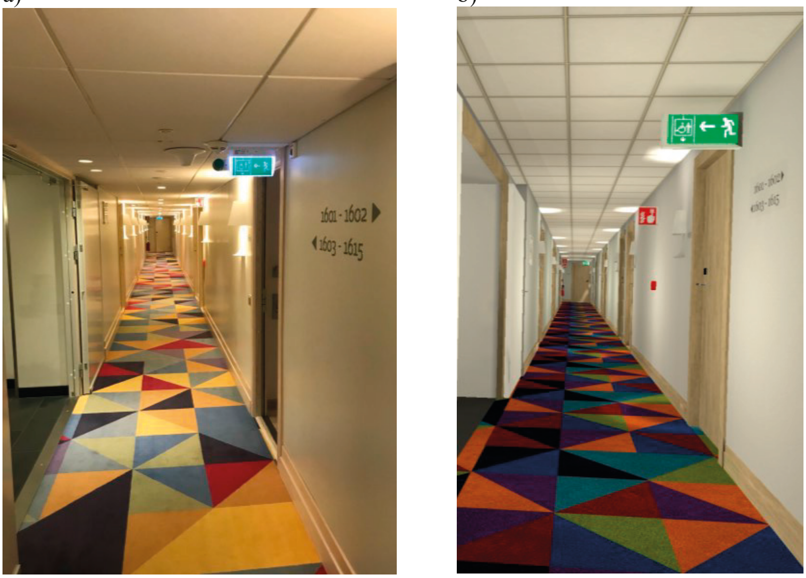

Figure 22 – Physical-world building used in the physical experiment run by Mossberg, Nilsson, and Andrée (2020) (a) and virtual environment used in Experiment 3 (b). Photograph by Axel Mossberg.

The behavioral data obtained in the different VR experiments presented here showed distributions similar to those obtained from the physical-world sources, which fulfills objective 2. In several cases, the similarities between the physical and the virtual samples were low. The virtual environments used were considered reasonably good representations of the physical-world environment they were meant to replicate. However, the behavior of participants in them showed there were differences that could have affected the data produced. These differences need to be considered (and if possible reduced) when running VR experiments. Nevertheless, the data was not only comparable, but also it showed similarities.

The results presented here represent the case of five different VR experiments with little in common other than the VR experiment method. They are not enough to draw strong conclusions on the validity of VR experiment method, and more studies are needed to assess the strengths and weaknesses of the method.

4.3. Limitations and considerations

The third and last objective of this study is to "identify limitations of the VR experiment method and considerations to be made in the pursuit of behavioral realism when using the VR method for collection of behavioral data". This section presents behaviors observed in participants that were unexpected in the experimental design. More than being just a peculiar finding, these behaviors highlight some less obvious limitations of the VR experiment method, or indicate aspects to be considered when designing a VR experiment. These limitations and considerations will be presented in the context of the corresponding experiments, and will then be used to discuss their effect on data produced by the VR experiment method.

The limitations and considerations are here presented in the context of the corresponding experiment in which its occurrence was clear. Since these limitations and considerations were unexpected, the experiments were not designed to measure or record them. They only became apparent after many participants showed the same behavior, sometimes even in different experiments.

4.3.1. Identified limitations of the VR experiment method

The VR experiments presented in this study attempted to replicate a fire event. However, some behaviors observed in participants in those experiments showed that VR, both as a technology and as an experience, has limitations on what it can replicate successfully. These limitations imply fundamental differences between VR and reality, which indicate cases in which the VR experiment method is not adequate to collect behavioral data.

4.3.1.1. Response to sounds in VR

In Experiment 1, the participant was in a bedroom in the second floor of the house when the fire started. In this experiment, it became apparent in the scenario with a working smoke alarm that some participants did not react to its sound (see *Paper I*). The sound was the typical high-pitch beeping of a domestic smoke alarm. There were no other sounds playing at the same time in the virtual environment that could mask this sound. Still, many participants in the smoke alarm scenario heard the sound and did not react to it. The expected reaction was to recognize the sound as a smoke alarm, or at least look around, trying to identify where the sound was coming from. The sound was directional, and its intensity increased the closer the participant got to the origin.

Participants who upon the alarm going off carried on with the same action they were performing before the alarm was triggered (or even casually picked up a different action) were counted as not reacting. Participants who interrupted the action they were performing by looking around them, trying to figure out where the sound was coming from, were counted as reacting. Around half of the participants did not react to the alarm, and no explanation was offered for that behavior. The sound was loud enough that it could be heard clearly, leaking out of participants' headphones, which means it was audible to the participant. This was a first indication that the sound of an alarm may not be enough to get the participant's attention in a VR experiment.

Experiment 4 presented another instance of participants ignoring auditory cues. In the particle accelerator tunnel, the alarm triggered during the bogus task was a loud siren. It was expected from the participants to react to it by at least looking around to see what that sound was or what was going on. Some did look around, some did not at first, but all participants walked away from their initial position, which could be their response to the alarm. The evacuation robot then reached them, displaying the evacuation direction (opposite to its own and the participant's travel direction). Since the robot moved at ceiling level, participants had to look up to see it. The robot emitted a light train sound as it moved on the rail. The sound, being directional, allowed participants to find its origin if they tried. However, pilot testing showed that it was hard to hear this sound on top of the loud siren, so the robot a ringing bell was added to the robot to make its presence clearer. The bell had a frequency of 1.5 Hz, making it clearly different to the sound of the siren.

After running the first robot scenario, it was observed that many participants saw the light projection on the floor indicating to them to turn back but did not look up to see where it was coming from or where the bell sound was coming from. Without looking up and seeing the robot, they also missed the screen it carried, which pointed to the same direction as the light projection. Since some participants did not turn back after seeing the light projection, it was believed that looking at the robot and seeing the screen it carried would make the message clearer. Therefore, in the second robot scenario, flashing green lights were added the robot, which easily caught the attention of participants. It is unclear why participants do not seem to react to sounds in VR as easily as they do to visual cues, but this experiment showed that auditory cues are less effective at catching participants' attention in VR than a visual cue, such as a flashing light. As a consequence, auditory cues in VR may not be as effective as those in a physical-world setup. As seen, additional visual features may be added to compensate for that difference, but these added visual features may not be credible in all fire events to be replicated in a VR experiment (e.g., a domestic smoke alarm is unlikely to be equipped with flashing lights).

4.3.1.2. Motivations in VR

Statistically significant differences were observed between the physical-world and VR samples in Experiment 2 in terms of the proportion of people performing a given action. As an example, around 90% of the physical-world fire victims in the hotel room broke the windows open to ventilate the room. In the VR experiment, the windows could not be broken but they could easily be opened instead (this feature was added for simplicity in the design of the virtual environment). Even though opening a window may be easier than breaking it, only 45% of the participants in the VR experiment opened the windows. This large difference could be explained by the motivations of each sample.

Participants in the VR experiment had no pressure to ventilate the room. The virtual smoke did not interfere with their breathing, it did not irritate their respiratory tract, and did not hurt them in any way. Victims in the physical hotel room were facing a life-or-death situation, and to their understanding at the time, breaking the window open may have been the only option they had to breathe fresh air. Therefore, the motivations the two samples had to deal with the smoke were very different. This case exemplifies the relevance of behavioral realism in a VR experiment. The motivations participants had to act in a given way in this virtual environment were nowhere near those of the physical-world fire. The VR experiment did not expose the participants to a scenario similar enough in terms of consequences, and therefore their behavior reflected that.

Something similar was observed during pilot tests in Experiment 1. Pilot testing showed how the lack of consequences in a VR experiment can lead to actions different to those observed in reality. *Paper I* describes how pilot testers kept trying to jump out of the window from the upper floor of the two-story house when they saw smoke coming up via the stairs. It is likely that in some physical-world fires the situation may be so dire that leaping out the window may be the best course of action. However, in both scenarios, as far as the participant could see, there was a large volume of smoke but no immediate threat. It was expected that participants would explore and try to assess the fire before making a drastic decision. Pilot testers showed their intent to jump out of the second floor without any further exploration of options. The decision to jump out of a second floor is much easier to make in VR than in real life. The informed consent form did not declare any risks of serious injuries in the VR experiment, and it is likely that participants did not expect to be hurt in any way during the experiment. This is why allowing the participants to jump out would overestimate the occurrence of this action. As a solution, and considering the relatively small fire participants would encounter if they went on to explore, no windows were openable or breakable in any way. While this feature may not be realistic for a regular single-family home, it removed the unlikely outcome of jumping out of the window at the sight of smoke. This is an example on how a virtual environment may need to be tweaked to maintain certain level of behavioral realism.

It should be noted that adding a feature like that does not affect the motivation of the participants, but only prevents a specific type of reaction. The motivations remain crucial for behavioral realism, since it cannot be expected that participants that have nothing to lose in a VR experiment can correctly reproduce behaviors of physical-world fire victims whose lives are at stake in an emergency.

4.3.1.3. Low engagement

In Experiment 2, some of the participants who did not try to stop the smoke coming into the hotel room through the vent, showed low engagement. It was observed that these participants had little to no interest to deal with the smoke. Some would not engage with it at all, and stood idle in the virtual hotel room as the smoke layer descended to the floor. This lack of engagement was emphasized, in some cases, by displays of irritation: sighing, shrugging their shoulders, shaking their heads. These participants seemed to not care about the emergency they were presented with, or would drag their feet to act. While the experiment was not designed to measure it, this lack of interest from the participants' side is especially remarkable as it hardly represents behavior observed in the physical world. People facing a fire may ignore the cues they get, but it is unlikely they would ignore it if the fire is perceived as a danger to their wellbeing. The fact that participants get to ignore the smoke, if they want to, highlights how low engagement can affect the behavior observed in a VR experiment.

4.3.1.4. Anachronism in fire safety design

Experiment 2 flagged the pitfalls of anachronism in VR experiments. Anachronism is here defined as chronological discrepancy, a misplacement of an object in time. In this case, the discrepancy lies between the expected fire safety features in a hotel room in modern times, and those available in 1980, at the time of the MGM Grand fire. Guests in that hotel had a TV and a radio, as most hotels did at the time. There were no smoke detectors and no sprinklers in the rooms, as per the fire safety strategy of the hotel, and there were no evacuation plans hanging from the inside of the room's door. The hotel followed the local regulations on fire safety, but the fire proved that the regulations were insufficient. Experiment 2 took place in 2018, and the fire safety features the MGM Grand lacked may be expected by modern participants. Some participants commented that they found odd there was no evacuation plan in the room, or that there were no sprinklers or alarms going off. These comments indicate that these participants' expectations were not met by the virtual hotel room they were in. They did not know, but the MGM Grand fire was pivotal in revamping the fire safety regulations for the state of Nevada at the time (Teague & Farr, 2009), requiring smoke detectors, sprinklers and evacuation plans, among other fire safety features (Guinn, 1981).

Experiment 5 also showed the effect of anachronism in the fire safety design of the nightclub. Similar to Experiment 2, some fire safety features were omitted in the virtual nightclub. There were no smoke detectors, no sprinklers, and no manual call-points (the timeline in the official investigation report does not include any form of alarm going off, either automatically or manually activated). At least one participant stated they roamed the nightclub in search of a manual call-point or a fire extinguisher but found none. Some participants mentioned that they thought it was strange the music did not stop when the fire was discovered, and that there was no alarm alerting people they should evacuate. Once again, participants used to modern features that are common in the present time cannot be expected to divest from their expectations and act as victims did in the past. It can be concluded that VR experiments may be more suited to replicate modern physical-world fires than those in the past, as the mindset of modern participants may not match that of the victims of the past.

4.3.1.5. Anachronism in technology

More than the state of the art on fire safety design, anachronism could also be hinted in the use of appliances. As mentioned before, the original hotel room in Experiment 2 was fitted with a TV and a radio, which were ubiquitous in hotel rooms in 1980. However, in 2018, radios are only seldom found in hotel rooms. Even if they are still found, the younger generation is more prone to use their cellphones to listen to the radio, if they even do so. The change in the times made a radio unnecessary in the room. The TV was kept, as it is normally found in hotels in modern times. Victims of the fire reported they listened to the radio or looked for information about the fire on the TV. The proportion of participants turning on the TV once the emergency started in Experiment 2 was much less than that of fire victims, as it can be seen on Figure 16. This difference can be due to several factors, anachronism being one of them. Younger generations are less reliant on the TV as main source of news (Shearer & Gottfried, 2017; Statens medieråd, 2019). Therefore, it is unlikely that they will reach for the TV in the same way victims of the fire did.

4.3.1.6. Waiting times in VR

Another factor that may have played a role in the decision to turn on the TV or not, at least in the case of the victims in the fire, is the time it took for them to be rescued. The victims answering the NFPA survey mentioned waiting up to four hours until rescue. This long waiting time could have led them to try to do more things, even if only to pass the time. The survey asked if they turned on the TV, not when or why. Participants in Experiment 2 were in the virtual environment for 15 min or less, giving them less time to think of possible alternative sources of information. There were several reasons to keep the time of participants in the virtual environment short. First, participants could not sit on the virtual furniture to wait comfortably for long periods. Moreover, the longer the time they spent in VR, the higher the risk was for them to experience motion-sickness. A third reason was to prevent them from asking for further instructions on what to do next. Therefore, it was decided to end the experiment when participants stood there waiting, not doing anything else. The difference between the participants, who knew they could interrupt the experiment at any time, and the physical-world victims, who knew they had no way out of the room, indicates a limitation of VR experiments in the reproduction of scenarios with long waiting times. This limitation, however, is not exclusive to the VR experiment method and also applies to other research methods, as participants signing up will like to know how much time it will take for them to participate.

Another instance of waiting times in VR can be observed in Experiment 3. That experiment replicated a physical experiment and a previous VR experiment in a CAVE. The CAVE experiment was designed to leave participants waiting for up to 20 min for the elevators to pick them up (Andrée et al., 2016). The data produced in the CAVE experiment showed that roughly 60% of the participants only waited 5 min or less before giving up on the elevator and using the emergency stairs. Therefore, in Experiment 3, the maximum waiting time was capped to five minutes. The results of Experiment 3 showed 60% of the participants leaving the elevator lobby in less than 2.5 min, and 90% of them had already left within 4.6 min. Participants in Experiment 3 seemed to be less willing to wait long for the elevators than those in the CAVE experiment. It is unclear why the difference in waiting times was so large.

4.3.2. Identified considerations to be made in VR experiments

The considerations presented in this subsection are different from the limitations presented before. These considerations may not fundamentally undermine the data collected, but may make the VR experience more realistic or more pleasant for the participant.

4.3.2.1. Anti-social behavior

Anti-social behavior was observed in Experiment 1. The non-player character simulating to be asleep on a couch became a victim of harassment or other forms of abuse from the participants. Some forms of socially unaccepted behavior were mild, like tickling the non-player character's feet or attempting to wake them up before the emergency started, for no clear reason. More concerning behavior was also observed, when participants threw heavy objects at the non-player character, or when a couple of them brought a knife from the kitchen and proceeded to stab the non-player character. None of these observed behaviors towards the non-player character would be expected in a physical experiment. Therefore, while VR can elicit realistic behavior, it is clear that not all behaviors observed in VR experiments are necessarily representative of reality.

Experiment 5 also included non-player characters, which were patrons in the nightclub. In this experiment, as mentioned in section 4.1.1.5, a made-up remote participant was mentioned to half of the participants in Experiment 5, with the intent of dissuading any proneness to anti-social behavior towards the non-player characters. Behavior that would not be acceptable in a physical-world setup was still observed: participants tried to make the non-player characters react, by grabbing their arms or touching their faces. The insistence of the participant trying to get attention from the non-player characters is understandable, as in a physical-world social environment, a small group of people would at least acknowledge the presence of a stranger stepping into the group. The non-player characters were not programmed to acknowledge the presence of the participant or to interact with them in anyway, which may explain why participants resorted to a more drastic approach to get their attention.

The conditions were also different from those in Experiment 1. In Experiment 1, the participant was in a quiet house with a single non-player character asleep on a couch. There was no activity in the house other than that from the participant. Therefore, they could dedicate their attention to trying to get a reaction from the non-player character. There were also many objects around the house that could be hurled to the non-player character. In contrast, in Experiment 5, there were not many objects around participants could use as a weapon to attack the non-player characters. Moreover, with the music, the disco lights, people dancing and talking to each other, participants could easily get distracted by their surroundings, dispersing their attention. Other differences between Experiment 1 and Experiment 5 are, of course, the type of building (a single-family home and a nightclub), the number of non-player characters present (a single one in Experiment 1, and around 150 in Experiment 5). The behavior was overall less violent in Experiment 5 than in Experiment 1, but this was valid also for the participant sample in Experiment 5 that was not told about a possible remote participant. Therefore, it is not possible to conclude that the addition of the made-up remote participant had the expected effect on social behavior in this experiment. This result, however, does not mean that made-up remote participants are not a valid way to try to enforce physical-world social norms, but rather that in this specific scenario they were not a good solution to test whether or not they work.

The stabbing would not have occurred if there was no knife in Experiment 1. The behavior of participants in Experiment 5 was less aggressive, but also there were not many objects they could weaponized. Therefore, it is worth to consider the objects to be included in a virtual environment and the effect they may have in the behavior of the participant. In some cases, like these two experiments, the anti-social behavior did not have an effect in the data collected (the behavior of participants towards non-player characters was not a behavior of interest). However, if the attitude towards other building occupants is relevant to a study, the anti-social behavior towards non-player characters should be taken into consideration in the experimental design.

4.3.2.2. Physiological reactions to the VR technology

The omnidirectional treadmill used in Experiment 1 was helpful for participants to navigate the entire virtual environment at their preferred walking speed, but it made several participants feel very ill, all of them being women. It is unclear what exactly made them feel sick, but for some of them the discomfort was too great to continue and the experiment had to be interrupted. Additionally, high level of perspiration was experienced by some participants, even among those who did not feel ill. It is unlikely that the perspiration was merely due to the physical effort made while using the omnidirectional treadmill. Even if participants would decide to move very fast, the distances within the house were too short for the amount of perspiration to be reasonable. Some participants were drenched in sweat, while others sweat just a little. The temperature and humidity levels in the experimental room was not measured, but it was within the limits of comfortable room temperature.

High perspiration was also observed in some participants in experiments that used the hand-controllers for navigation (Experiments 2, 4 and 5). These cases were fewer than in the case of the omnidirectional treadmill, but were not recorded as they were unrelated to the data that was collected in those experiments. Nevertheless, their occurrence was noticed. Some participants handed back the hand-controllers drenched in sweat. In these experiments, it was also observed that a small portion of participants tended to lose balance while they were in the virtual environment, even though they were not moving physically but just standing in the same position. Again, this was not recorded as it was out of the scope of the data to be collected, but the occurrence is reported here.

Neither the perspiration nor the problems keeping balance seem to have been detrimental to the data collected in the experiments, but comfort for the participant should also be considered. Uncomfortable participants may have an incentive to wrap up the experiment quickly, limiting the number of actions they would have performed if they were comfortable.

4.3.2.3. Implications of the VR equipment

The level of dexterity needed to block the vent in the virtual hotel room Experiment 2 appeared to be too high. The vent was placed on a wall on the skirting level, to make it accessible for participants and to simplify its blockage. In the physical hotel room, the vent was placed at ceiling level, forcing the guests to climb on furniture to reach it. Climbing virtual furniture in a scenario with room-scale based locomotion is not possible without providing physical pieces of furniture that can be tracked by the VR equipment. Moreover, victims of the fire had to fit the materials they used (bedding, towels, etc.) in the vent, compressing them to fill the hole tightly. Compressing objects in a virtual environment is not only a difficult visual feature, but it is also extremely costly in terms of computational power. Due to this difficulty, the vent was placed near the floor and any object close enough to it (around 5 cm from it) would have an effect on its smoke plume.

Initially, the entirety of the vent needed to be covered, as any uncovered parts would still allow smoke into the room (as it would be the case in reality). The mass flow rate of the plume was inversely proportional to the area of coverage of the vent. However, during pilot testing, it became apparent that participants did not recognize the reduction of the mass flow rate of the plume, and assuming it was not possible to block the smoke at all, they gave up. In other words, pilot testers did not seem to care about the parts of the vent that remained uncovered. It is hard to imagine adults in a physical-world environment not realizing that they need to cover the entire hole of the vent for the smoke to stop flowing out of it. Nevertheless, as a solution, the vent was simplified to such a degree that even partial coverage of it reduced the mass flow rate (drastically if the object was dry, and completely if it was wet). This simplification proved to be enough for participants to understand their effort was fruitful and it was in fact possible to stop the smoke. This case showed that physical actions do not translate seamlessly in VR, and certain concessions are needed to ensure participants understand the mechanics of the virtual environment.

The HMD used in the experiments has a nominal field of view of 100° in the horizontal, which is a drastic reduction from roughly 180° natural human field of view (Zihl, 2006). This implies a reduction in the peripheral vision, limiting the visual cues participants can receive from their surroundings. The reduced peripheral vision is obviously a difference between VR and reality, but its impact on the data was not measured. It is possible that the reduced field of view may only make the VR data more conservative. If a participant can perceive a visual cue in VR despite the reduced field of view, it is likely that they will perceive the same cue quicker had their field of view been wider. Nevertheless, this difference in field of view is inherent to the type of equipment used, and future VR equipment may offer wider view angles.

4.3.2.4. Collision avoidance and behavioral realism (2)

Experiment 2 was conducted in a large physical room, in which a 5 m x 5 m empty area was demarcated on the floor. Participants were told that was the space in which the experiment would take place and that they would be wearing a wireless solution that would allow them to walk freely in it. They were also told that they could feel safe as no furnishings were protruding into the demarcated area, and a buffer zone around it would prevent them of bumping into furnishings by accident. Once they wore the HMD, they could not see the physical surroundings, only the virtual ones, and they behave accordingly. To move around in the hotel room, participants walked in the physical space avoiding the areas where the virtual furniture was placed. To walk into the bathroom, participants used the bathroom door. No participants walked through the virtual furniture or through the virtual walls. They respected the virtual obstacles and moved around them as they would do in the case of a physical one, even though they knew there were no obstacles in the physical room they were in. The collision avoidance observed in this VR experiment evokes the results of the behavioral realism study conducted by Kisker et al. (2019).While this collision avoidance was observed in all five VR experiments performed in the context of this research work, it becomes remarkable in Experiment 2, as there were no restrictions that could prevent them from walking through the obstacles. This result highlights the fact that even though participants may be very much aware of the lack of real risks in a VR experiment, the VR technology can trick the human brain into processing the virtual environment as a physical one and into reacting accordingly to some extent.

4.3.2.5. Participants bring their own expectations

In Experiment 3, it was remarkable that some participants said they thought there was no emergency staircase, which influenced their choice of using the elevators. The Swedish building code does not contemplate the possibility of high-rise buildings without any staircase (BFS 2011:6, 2011). It is not expected that participants know the local building codes, but it is unclear why they thought there may be no stair. This shows that participants bring their own expectations to the VR experience. If they do not expect a given feature or a given action to be possible, their behavior in the virtual environment will reflect that. Nevertheless, even if some participants truly believed there were no stairs, that changed at some point while waiting for the elevators, as only a handful of participants kept waiting until the maximum waiting time of five minutes was reached.

4.3.3. Fulfillment of objective 3

The third and last objective of this research work expected to "identify limitations and considerations to be made when using the VR experiment method for collection of behavioral data". Several limitations and considerations were presented and discussed. They were many, and they varied across the different VR experiments performed. It is clear that they are not necessarily applicable to all possible VR experiments, and that some can be hard to identify before running pilot tests or the proper VR experiment. There is still much to learn in terms of limitations and considerations of the VR experiment method. In some cases, features can be added to reduce the effect of the limitations, but if that is not possible, the VR experiment method may not be suitable for that particular study.

The results presented in this research work show several limitations of the VR experiment method, as well as some considerations on how to make a good use of the VR technology to collect behavioral data. These results are mostly based on experience acquired by observation of the behavior of the hundreds of participants included in the different experiments and pilot tests. More than a study on human behavior in fire, the experience acquired in this research work provided an insight to human behavior in VR, which needs to be understood to in order to produce realistic VR experiments. Participants showed the many ways a situation can be interpreted and the many reactions that it can trigger. There are countless anecdotes of remarkable behavior observed in participants and others. These qualitative observations may not constitute a rule, but each add a new possible outcome to those already expected in the experimental design. They showed that some features can cause frustration (e.g., operating swinging doors), how they could succeed at bending the rules of the virtual environment (e.g., sticking their heads through virtual walls to see the other side), how sounds can be interpreted in VR (e.g., "that was the smoke alarm? I thought that was a microwave"), and to which extent VR can trick participants into believing the virtual surroundings are physical (e.g., attempting to put their weight on the virtual furniture). Some of these may only have been observed in a handful of participants, they showed how participants can behave in unexpected myriad of ways. Future virtual environments can take these observations into consideration and improve their design.

There are some mundane reasons why participants may behave in an unexpected way (i.e., curiosity about the VR technology, misinterpretation of auditive cues, trouble getting used to the hand-controllers), which can be compensated with relatively simple solutions (i.e., giving them more time to explore and get used to the equipment, running the experiments in a quiet area). However, there is a key physical-world component that can be hard to incorporate in a virtual environment: a social context.

Humans are highly social beings, and much of the human behavior follows a code of conduct (Lapinski & Rimal, 2006). There are many examples of the application of the code of conduct, but a few basic ones are highlighted here: being polite, performing one's role in a given circumstance, adjusting the tone of the voice to the audience, or the volume to the ambient noise, etc. These are very basic, and they emerge naturally among most adults. During a virtual experience, however, participants seemed prone to disregard social conventions. A possible explanation is that they know no real harm is caused in a virtual environment. Participants know that any damage done in the virtual environment will vanish the moment the simulation is restarted. They know there will not be repercussions for defacing a virtual room, no bill will be sent to them to refurbish it, the police will not be called for assaulting a non-player character, nor will any lawsuits follow. Therefore, the participant may see no reason to refrain from throwing furniture out of the window or see no incentive to be civil to a non-player character.

Those behaviors may be extreme, and more of an exception than a rule among the participants who joined the five experiments presented here. However, following the same line of thought, other more behaviors of interest can be similarly affected: why would a participant care about the well-being of a non-player character? The non-player character is not real, it cannot die. Why would the participant care about extinguishing the fire? It is not their property, and no real damage is done as the simulation can be restarted. Why would the participant care about evacuating safely? The informed consent form they signed stated clearly the risks related to the experiment, and none of them represents a serious risk to their health. With little incentives for them to behave as they would do in real life, the behavioral data produced in VR experiments has limitations.

At this stage, it is not possible to single out the kind of behavioral data VR is best at predicting. However, the motivations can provide a hint. As it was observed in Experiment 2, around half of the participants did not block the vent to prevent the smoke from entering. In contrast, 90% of the physical-world victims did so. The participants did not have the same motivations the victims had, as the participants had much less at stake. In comparison, in Experiment 3 the proportion of people using the elevator as their first choice did not differ much between the VR sample and the physical sample. The motivations of participants in these samples were much more similar. The consequences for both groups of participants were about the same, as none saw any immediate threats to their well-being. In hindsight, it is not surprising that similar motivations produce a similar outcome. Unless the motivations can be equalized in some way, the VR experiment method may be better suited for cases in which the consequences for the participants are not too far from those expected in a fire event. Examples of these cases are system design, and way-finding experiments.

If life-threatening conditions are to be studied, a deeper look into the inclusion of consequences into the VR experience is needed. It is not possible to suggest a good way to introduce those consequences without testing them first, as none of these five experiments attempted to do so. The effect of this lack of consequences was identified after running Experiment 2, and some ideas are presented in the Future research chapter (see Chapter 6). Nevertheless, it is here emphasized that one of the main advantages of the VR experiment method, the lack of risks, is a double-edged sword – participants are aware nothing can harm them in VR and some act accordingly.

4.4. Assessing the VR experience

The results and discussions presented in this section refer to the assessment of the VR experience by the participants. They are not directly related to a specific objective but they add depth to the results presented in the previous sections. The assessment made by the participants illustrates the perception they had of the virtual environment and their VR experience, and helps to understand the participants' mindset, and therefore their actions. If the virtual environment is not seen as realistic or if operating in it was too difficult, the quality of the collected data may be affected.

The assessment was included in the questionnaire participants filled up after finishing their VR experience. They were asked to rate different features of the virtual environment and their VR experience using a Likert scale. The results of their assessment are presented and discussed in the following subsections.

It is worth pointing out that each participant experienced a single run of a single scenario in a single experiment. This means that the ratings that are presented here are not an assessment between experiments, but one within each of them. The same is valid for the assessment of the type of locomotion used, which was not the same for all five experiments.

4.4.1. Results on the assessment of the VR experience

In the following subsections, the results from the assessment performed by the participants will be presented in terms of realism, presence, sensations, navigation, and use of the hand-controllers.

4.4.1.1. Realism

Participants in each experiment were asked to rate the realism of the VR experience, in order to assess the overall credibility of the experiment. The rating ranged from low (1) to high (7). Each experiment included different aspects of the virtual environment to be assessed in terms of their realism. Figure 23 presents the proportion of participants in each experiment giving each of the possible ratings to realism of the visual components in the virtual environment they were exposed to. The question on realism of the visual component was specific for each experiment as the virtual environments used in them had very little in common. Examples of the visual components they were asked to rate are the smoke, the fire, the hotel room, the elevator lobby, the tunnel, the crowd in the nightclub, etc. As shown on Figure 23, most participants in each experiment gave medium to high ratings for realism.

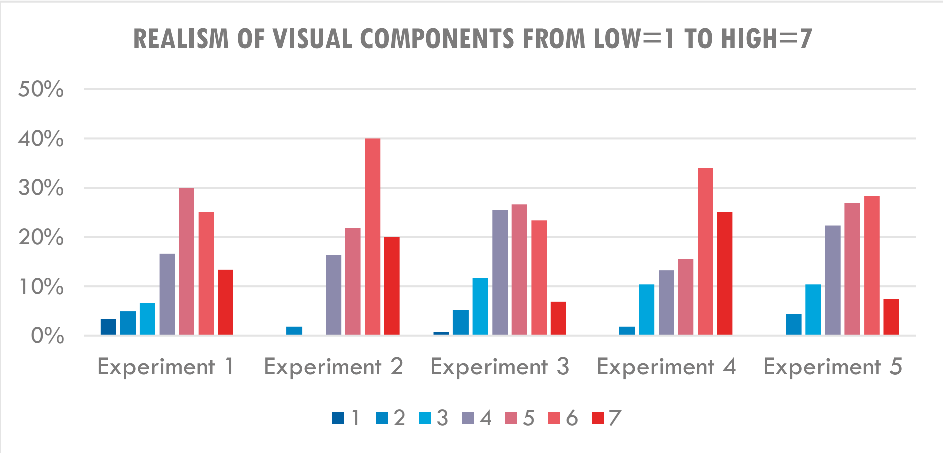

Figure 23 – Proportion of participants per experiment giving each of the possible ratings for realism of the visual components in the corresponding virtual environments

4.4.1.2. Presence

The level of presence experienced by the participant was deemed relevant to understand whether they felt part of the virtual scenario around them. The question in the questionnaire asked “Did you feel immersed? (did you ‘forget’ that you were in a laboratory instead of a hotel room?)”, with small variations to include the virtual environment used in that experiment. It should be noted that the question was phrased using the word “immersed” instead of “present”. Given the definitions for the term immersion and presence introduced in Chapter 2, the question is not formulated correctly, and it should read “did you feel present?”. However, the added explanation in parenthesis may have been more relevant to the participant than the term used. That explanation is still aligned with the definition given for presence in this research.

The question was a multiple choice one, offering “yes”, “no”, and “other” as possible answers, with “other” allowing participants to explain their view using their own words. Most participants who chose “other” gave descriptions of a mixed state, ranging from experiencing being in both at the same time, to only once the emergency started”. Figure 24 presents the answers given by participants in experiments 1, 2, 3 and 4. In Experiment 5, an attempt was made to offer a more nuanced set of answers to this question. Instead of the “yes”, “no” and “other” options, a scale was offered from 1 “a little” to 10 “a lot”. The answers provided are presented in Figure 25.

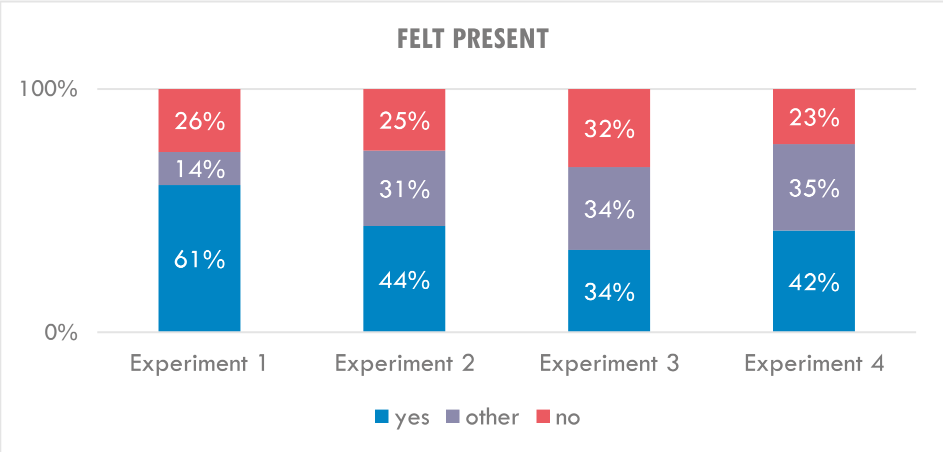

Figure 24 - Participants' assessment of feeling present in the virtual environment in experiments 1 to 4. Data from Experiment 5 is not included in this figure.

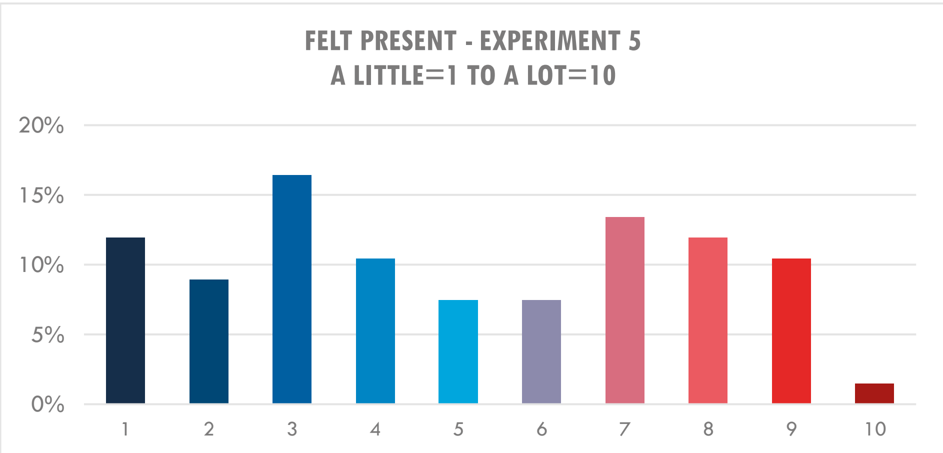

Figure 25 – Proportion of participants in Experiment 5 giving each of the available ratings.

As seen in Figure 24, there was a higher frequency of "yes" answers except in Experiment 3. However, it is not possible to tell if the answers provided were biased due to the fact that selecting "yes" or "no" required less effort than to describe the subtleties of the in-between state. Figure 25 shows how that could be the case, as only 1% of the participants in Experiment 5 gave the highest rating, and the distribution of answers does not reflect, even remotely, the answers of participants in the other four experiments. It is possible that Experiment 5 offered a much lower level of presence than the other four experiments, but such an extreme difference would be unlikely. More importantly, the difference between the two types of answers to the same question highlights how participants' answers can be affected by the design of the questionnaire.

4.4.1.3. Sensations

Participants were asked to rate some sensations they felt while in the VR experience. The three sensations, insecurity, stress and fear, were rated from 1 (low) to 7 (high). Their assessment was intended as an insight into how emotionally demanding the experiments were. Figure 26 presents the proportion of the total poll of participants across all five experiments giving each rating to the different sensations. As it can be seen, less than 40% of the participants gave the three highest ratings, indicating the experiments were not very emotionally strenuous.

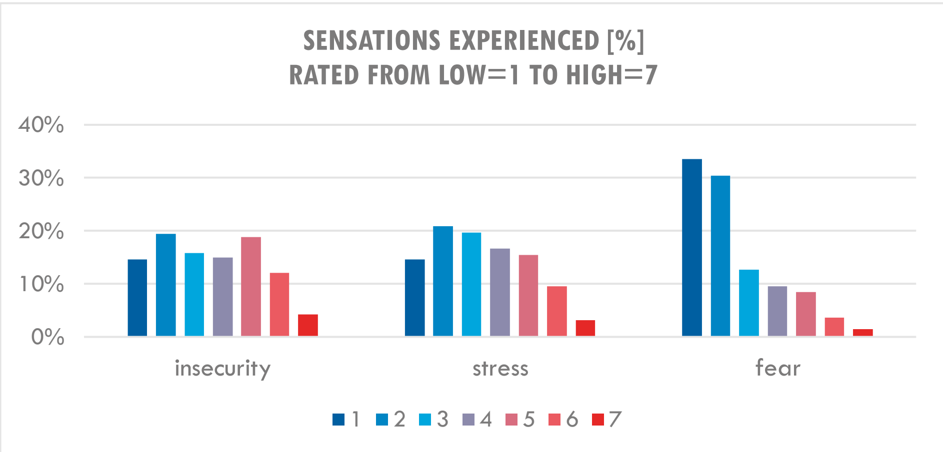

Figure 26 – Sensations rated by participants across all five experiments. The data is presented in percentage of the total number of participants giving each possible rating

4.4.1.4. Navigation

In the questionnaires, participants were asked to assess how easy it was for them to navigate the virtual environment and to use the hand-controllers. The type of locomotion adopted was based on the conditions of each experiment. In Experiment 1, an omnidirectional treadmill was used, making the type of locomotion motion-based. In Experiment 2, a wireless solution was used, so the locomotion type was room-scale-based. In Experiment 3-5, hand-controllers were used, by which the locomotion type was controller-based. Participants were asked to rate the level of difficulty of navigating the virtual environment they were exposed to. Figure 27 presents the proportion of participants giving each of the available ratings, from easy (1) to hard (7).

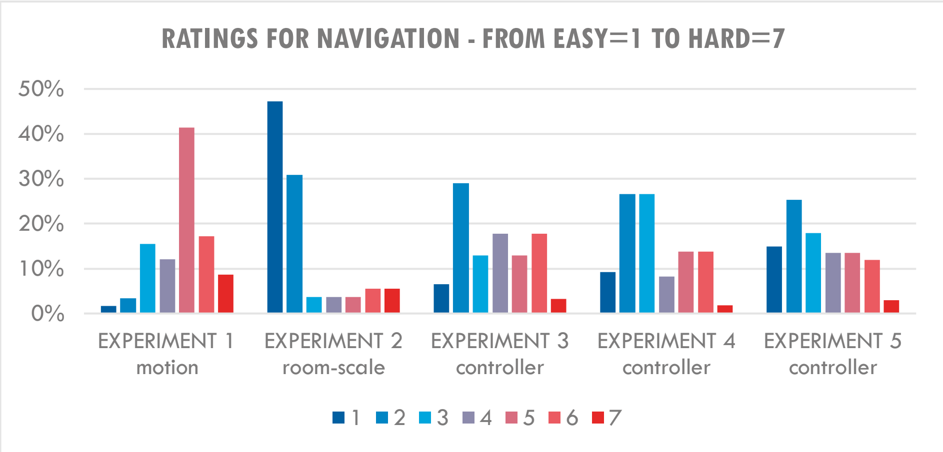

Figure 27 – Proportion of participants giving each rating in each experiment to the difficulty of navigating the virtual environment. The labels on the horizontal axis indicate the experiment, and the type of locomotion it was based on

Experiment 1 was the hardest one to navigate, which came as no surprise because many participants in that experiment expressed that it was difficult for them to use the omnidirectional treadmill. The room-scale locomotion used in Experiment 2 received the best ratings, with almost 80% of the participants giving the two lowest values.

4.4.1.5. Use of hand-controllers

As mentioned before, all participants used the same equipment (HTC Vive series HMD and hand-controllers). The hand-controllers have a trigger button and a touchpad. Both had functions assigned depending on the experiment. The trigger button was used to hold virtual objects in the participants' hands. As long as the participant held the trigger tight, the object would remain in their grip. By releasing the trigger they released the object. All experiments used the trigger as a grip function.

The touchpad was used to command movement in the experiments with controller-based locomotion. By touching the touchpad, the participant could move forward in whichever direction they were looking. Additionally, the touchpad could also be pressed down as a button, which was used to enact an additional function. This button function of the touchpad was used for objects like the remote controller for the TV or the extinguisher, when the participant had to hold the object in the hand to operate it (e.g., using the trigger to hold the extinguisher, and pressing the touchpad to discharge it). Experiments 1 and 2 benefited from that additional function, since the touchpad was not needed for navigation. However, in Experiments 3 to 5, adding that functionality was deemed too complex, since participants were using the touchpad already for navigation. Given the relatively short time participants have to learn how to use the hand-controllers, it was important to keep their use as simple as possible. Therefore, in Experiments 3 to 5, which relied on controller-based locomotion, participants learned how to use the touchpad for navigation, and the use of the trigger button (one button). In the other two experiments, participants learned how to use the touchpad as a button, and the trigger (two buttons).

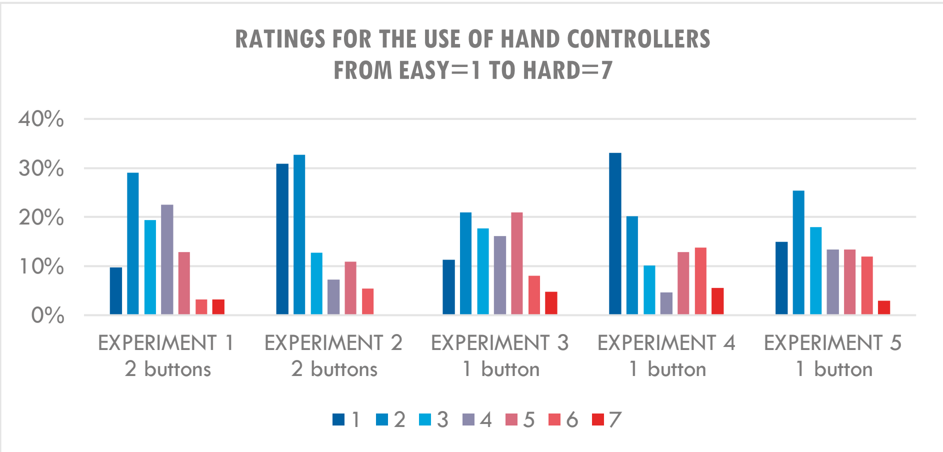

Figure 28 – Proportion of participants giving each rating for the difficulty of using the hand-controllers in each experiment. The labels in the horizontal axis indicate the experiment, the number of buttons participants were trained to use, and the average rating in parenthesis

The use of the hand-controllers was assessed by the participants by giving a rating between easy (1) and hard (7). All participants used the same hand controllers, and no mention was made to them about the number of buttons they used. Figure 28 presents the proportion of participants giving each rating in the different experiments, indicating the number of buttons they used. The differentiation between the numbers of buttons as presented in Figure 28 is made to give a sense for the different levels of complexity on the use of the hand-controllers. As shown in Figure 28, participants tended to give ratings on the easy side of the scale, with Experiment 2 being again the one with the best ratings.

A high level of difficulty in the use of the hand-controllers would have been detrimental to the behavioral realism in VR experiments, as participants struggling to perform simple actions would mean a difference with a physical-world experience. The medium to low ratings given by participants indicate the struggle was not too severe to impede normal operation of the objects they found.

4.4.2. Implications of the assessment

As seen on Section 4.4.1, participants gave an overall positive assessment of the perception of realism and their sense of presence. The sensations they rated were in general not extreme. The stress level experienced was overall moderate, although a higher level of stress would be desirable for a better agreement between the motivation of the participants and those of physical-world victims. Nevertheless, no peaks of stress were observed, which means that the emotional burden was not too high for the participants in these experiments.

The assessment of the equipment showed the benefits of using room-scale locomotion, as participants in the experiment with this type of locomotion (Experiment 2) gave the lowest rating for difficulty. The controller-based locomotion (using the touchpad on the hand-controllers) was assessed on an overall medium scale of difficulty. The motion-based locomotion used (the omnidirectional treadmill) was overall assessed as harder to use. As mentioned before, participants used a single type of locomotion, and their assessment was not relative to the other available types. However, participants using the omnidirectional treadmill felt motion-sickness symptoms much more frequently and more intensely than participants using the other two types of locomotion. Moreover, these participants did not have much time to learn or improve the technique on how to use the omnidirectional treadmill, since there was only enough time for them to do a basic training before the experiment. It is possible that, had they had more time, they could have gotten better at using it. The omnidirectional treadmill was used constantly during the development of the virtual environment without experiencing any symptoms of motion-sickness by the researchers.

The ratings given on the use of hand-controllers was overall homogeneous, with a level of difficulty from easy to medium on average. With only two buttons to be used, the hand-controllers were not too complex, which is an advantage when there is only a couple of minutes for participants to get used to them. However, their simplicity limited the functionality of the participants' hands in the virtual environment. These hand-controllers gave participants the faculty of grabbing virtual objects, and in some cases, operate them with the button function of the touchpad (e.g., to change channels on the TV using the TV remote-controller). These two functions are analogous to closing a hand into a fist and the ability to exert pressure with the thumb or the fingers. The experiments were designed to work with only those two functions, but human hands can perform many more. Digital dexterity was severely reduced. A simple task like dialing on a smartphone became a technical challenge, as it can be cumbersome to press a number on a dial pad when using a fist instead of a thumb. Even the button function of the touchpad could not help with the dial pad, because it is a single button to press, and the dial pad needed many more buttons. By the time the last experimental study was near completion, a new VR compatible hand-controller was launched, which enables some individual tracking of fingers. This hand-controller would have been more intuitive for the participants in some use cases. Other solutions that would allow for finer movement of the fingers did exist at the time, such as haptic gloves and optical hand-tracking devices. Unfortunately, they were in relatively immature states of development and not suitable for experiments. It is likely these devices will become easier to implement and use in the future.

4.5. The research strategy used

The research strategy adopted intended to discern the validity of behavioral data produced using VR experiments. It was thought that a quantitative comparison between VR data and physical-world data was an objective way to determine how valid the VR data was. However, from the beginning of the design phase for Experiment 1, it was clear that a virtual environment does not replicate reality by default. The virtual environment needs to elicit realistic behavior, which is not an inherent property of said virtual environment. This first realization was key to understand that the virtual environment can only be as good as the input it is based on. Much time and craft is needed to make it *appear* realistic. More than just the realism of the graphics, large effort was put into making the emergency seem credible, cohesive and compelling. This effort, when successful, went unnoticed. Failures, on the other hand, were crystal clear when the virtual environment did not respond as the participant expected.

Having found out how hard (and sometimes impossible) it was to program the virtual environment to simulate certain physical-world conditions, the second realization was that *if* the virtual environment cannot replicate reality well, neither will the data obtained. Therefore, it became paramount to understand what can and cannot be achieved in a modern virtual environment, given the constrictions in terms of time and equipment.

The first realization indicated that physical-world behavior is not inherent to the VR experience, or at least not by default. For example, in Experiment 2 half the participants did not put enough effort to block the smoke from coming into their hotel room. The same applies to Experiment 1, in which the sleeping non-player character was a victim of different kinds of aggression. These are just two of the many examples of the shortcomings of the VR experiences failing at reproducing a fire event realistically. Therefore, instead of assessing the validity of the data, the focus of this research is on improving the realism of virtual environments for VR experiments.

The research strategy applied is reasonable, as the physical-world data was used to gauge the aptitudes of VR experiments at recreating it. Differences between the physical-world and the VR samples highlighted possible sources of bias in the VR data. Some of those sources of bias could be reduced in some cases, and in other pointed out limitations of the VR experiment method. There is no standard procedure to ensure the virtual environment is realistic without pilot tests and cycles of trial and error. Even then, what is deemed realistic can also be subjective. Therefore, the only input that could be used to check the realism of the VR experience is the comparison of results with physical-world data. To that extent, the research strategy adopted was not only appropriate but also effective.

4.6. Reflection on statistical analysis

As explained in section 3.2, the p-values were used to assess how similar the two samples (VR and physical-world) were in each experiment. If the p-values were consistently high or low, it would have been easy to draw a conclusion on the overall ability of VR to replicate physical-world behavior. However, the p-values obtained in the different analyses were distributed between 0 and 1, with no clear trend towards a specific value. This wide spread of results can be explained by the samples used and the understanding of behavior in VR.

4.6.1. Sample issues

Sample size is a known issue in experiments with human participants, as mentioned on section 4.2.2. Constraints of the budget, schedule conflicts and overall insufficient number of volunteers force researchers to reconsider their expectations for the ideal sample size. Moreover, some participants signed up for one of the five experiments one year, and tried to take part in one a year or two later. People who had taken part in a previous experiment were rejected in any subsequent experiment to avoid any learning effects. Unfortunately, this screening impacted the sample size. The limitations in the recruitment and the resources available for each study meant that the sample sizes were set at a minimum of 30 participants per scenario, but that was not always achievable. Therefore, convenient sampling was necessary. With this relatively small sample size, a handful of participants can largely alter the results of any significance testing, as mentioned on section 4.2.2.

The relatively small sample size was still large enough not to violate the assumptions of Fisher's exact test, which is the test that was used in all cases. However, this test may not be ideal given the aim to assess similarities. Different suitable statistical tests were considered as an alternative. Some of those tests were not applicable due to the reduced sample size, and some due to the reduced number of possible rankings (yes or no answers being only two rankings while the test needed at least five). Confidence intervals could not be calculated because the samples were too small. The issue with sample sizes, however, is not inherent to the VR experiment method, but in general to any studies relying on unbiased participants.

Convenient sampling has further implications, such as the composition of the sample. The VR samples diverged from the physical-world samples in many ways. The demographic of the physical-world samples was not necessarily matched by the VR samples. The VR samples were conformed mostly by students at Lund University in three out of five experiments, while the physical-world samples were not. People volunteering as participants had an interest in VR, which may not have been the case in the physical-world sample. In some cases, the physical sample belonged to a different geographical location, like a much larger city or a different country. In other cases, the physical sample belonged to a different point in time (e.g., the MGM Grand fire took place in 1980).

These issues with the samples are not negligible and may have affected the results. A clear example is the case of anachronism discussed in section 4.3.

4.6.2. Understanding behavior in VR

The five experiments included in this study initially intended to simply compare the VR and the physical-world samples and validate VR experiments as a research method, assuming the realistic aspect of the virtual environments were all participants needed to behave as they would do in a physical-world environment. The behavior of participants in the different experiments proved that assumption wrong, or at least incomplete. Aware that their well-being was not at risk, some participants engaged in behaviors that are unlikely in a physical-world case. Consequently, the data produced would not necessarily match the physical-world data. The differences between the virtual environments and the physical-world fire event (detailed in the limitations and considerations presented in section 4.3) were so large, it was evident the two samples behaved in different ways. The response to sounds, the differences in motivations, the issues with low engagement, the difficulties operating the VR equipment, are some of the many aspects that highlighted the differences between virtual environments and the physical-world ones.

The present research work offers a meaningful contribution to the development of the VR experiment method by exposing those limitations and considerations relevant when applying the method. Despite the large variation in the results of the statistical analysis, the findings presented here are an early step towards the implementation of a protocol for the application of the VR experiment method as a research tool for Human Behavior in Fire research.

5. Conclusion

This thesis aimed to study in detail how to apply Virtual Reality in Human Behavior in Fire experiments to collect behavioral data. Five VR experiments were conducted, replicating well-documented physical-world events. The data collected from the VR experiments was then compared to the physical-world data to assess how similar the two datasets were. The major contributions of this thesis are listed as follows:

- The VR data matched the physical-world data in several instances, showing an agreement in the data of both datasets. Behavioral trends were observed in VR in a similar way as in a physical-world context. The behavioral data showed distributions similar to those from the physical-world sources.
- Although the replication was not exact, the differences highlighted limitations of VR experiments or aspects that need to be improved in the future. These limitations and considerations, included in this thesis, make up a simple guide for researchers in Human Behavior in Fire to avoid pitfalls, and advance the VR experiment method.
- Motivations have a large impact in the participants' behavior in VR. Human behavior does not occur in a vacuum, but it is a function of the environment. The motivations that guide the behavior of fire victims need to be artificially introduced in a virtual environment to adequately reproduce their behavior. Without the same motivations, participants will not have reasons to behave in the same way. Future studies can indicate possible ways to introduce consequences to the participants' VR self to approximate their perception of risk to that of fire victims.
- The code of conduct that dictates human behavior in a physical environment needs to be adequately represented in a virtual environment to promote behavioral realism. An inconsistency can lead to participants showing different behavior in VR, which is specially notorious in social scenarios in VR. The occurrence of anti-social attitudes towards non-player characters could be reduced by an artificial enforcement of social norms in VR. Further studies are needed to identify solutions to artificially enforce social norms in VR.

- Virtual Reality is clearly capable of providing a sense of realism to the participant's experience, but work is needed to compensate for the differences between VR and reality. The incorporation of additional stimuli by an incursion into Augmented Reality could enhance the perception of realism for the participant, but more studies are needed to determine the size of their effect.
- A faithful reproduction of reality may be unlikely for the state-of-the-art of the VR technology. An imperfect reproduction, however, is very much possible with enough expertise in VR. Expertise in the VR experiment method implies a deep understanding of the limitations of the VR technology and how they can affect participants' behavior.

6. Future research

As brought up in the discussion of the different results in the scope of this research work, future studies could improve the realism of the VR experience for the participant and by extension the quality of the behavioral data collected. Some shortcomings of the five experiments presented are due to an insufficient understanding of the limitations of VR reproducing reality. Those limitations have a noticeable effect on the behavior of the participants in a VR experiment for collection of behavioral data, and therefore on the data itself. The future research proposed here takes a step back from the fire evacuation objectives, and focuses on improving the realism of the VR experience for the participant. The proposed studies explore the nature of human behavior in VR. Their outcomes would have a positive impact in the realism of the VR experience and therefore the quality of the data of VR experiments for data collection in human behavior in fire.

6.1. Improving realism in VR

The five experiments included in this research work relied in different features to enhance the realism of the VR experience for the participants. However, the experiments were not designed to study the effectiveness of those features or whether there were better solutions. Therefore, the studies presented in the following subsections aim at trying different solutions to determine their usefulness. The realism of the VR experience can be improved in two fronts: the behavioral realism and the sensorial realism.

6.1.1. Behavioral realism

Improving the behavioral realism refers to the features of the virtual environment that will approximate the behavior of the participant in it to that expected in the physical world. As mentioned before, without consequences to the participants' actions, the motivations guiding their behavior will not be the same as in the physical world. Studies need to assess how to implement consequences to the participants' actions during the VR experience. Physical-world consequences could be of different kinds: injuries or physical damage, social scorn, liability for one's actions (civil lawsuits), and even law enforcement (criminal justice). It is obvious that participants cannot be served with lawsuits or be prosecuted for a behavior in a virtual environment that is not translated to the physical world (e.g., stabbing of a non-player character). However, there may be alternatives to simulate physical damage and some kind of social backlash for their misbehavior in VR.

Simulating physical damage

Physical damage is the most obvious consequence that can hardly be simulated in VR, even in the unlikely case the ethical considerations deem the experiment acceptable. The infliction of pain may not be necessary, if participants become aware that their virtual wellbeing is suffering. *Paper II* suggests the possibility of using a health bar, similar to those used in video games. The health bar could be tested in a range of configurations, from simple system to a complex one. A simple system could be a plain a green bar that is shortened by close proximity to flames, or by being within the smoke layer. A more complex health bar will indicate what is causing the damage (e.g., pain, irritation of the mucous membranes, asphyxiation) and the rate of damage being done can vary accordingly (e.g., burns cause large damage quickly, while exposure to toxic gases takes a longer time to cause as much damage). An irritation of the eyes could be indicated by adding a blinking effect that simulates the closing of the eyelids by blacking out the screens momentarily. A vignette can also be added to imply pain or damage. A study could also test if the sound of an accelerated heartbeat, added to signify agitation, increases the levels of stress in the participant.

Social response

Experiment 5 included a scenario in which the participant was told a remote participant may or may not join during the VR experience. It is possible that the false remote participant had no effect enforcing good behavior in the real participant, especially since participants did not have either the time, the means or the interest to cause significant damage, compared to the behavior of those in Experiment 1. It could be argued that opportunity makes the thief, and participants in Experiment 1 would not have stabbed the sleeping non-player characters had they not had access to a knife. However, some participants threw various objects at the non-player character, which highly inappropriate even if non-lethal. This behavior is also highly unlikely to be observed in a physical-world experiment. The point of correcting the behavior of the participant towards non-player characters is not policing their behavior, but rather setting up a context for their actions similar to the physical world. The aggressive behavior towards the non-player character is a reminder that not all behaviors in a virtual environment can be extrapolated to reality.

There are several ways social norms could be enforced, but studies need to be performed to determine their effectiveness. The remote participant, whether real or not, could be a good solution in other scenarios. For example, one in which there is a verbal exchange between the participant and another character in the virtual environment. Other ways to explore the response to a non-player character could be to have the non-player character resembling another person in the experimental room. That other person could be the researcher, another participant, or a confederate pretending to be another participant. In any case, if the participant can recognize the similarities between the non-player character and the real person, they could act in a more caring way towards the non-player character.

6.1.2. Sensorial realism

The VR experiences provided by the five experiments were heavily reliant in visual input. Therefore, the effect of the outlook of the virtual environment needs to be considered. Other senses, like the sense of touch and the sense of smell, could be added to the VR experience to provide a more complete experience.

Visual realism

The effect of visual realism in VR can also be studied further. The virtual environments used in the five VR experiments included in this research were far from being photorealistic. Some of them were better than others in terms of lighting and the application of the right textures to materials. It is not possible to tell if the difference in the level of skills at producing a visually realistic virtual environment affected the data produced. However, it can be determined whether high levels of visual realism have a positive effect in the presence experienced by the participant. A study can be conducted in which participants are asked to assess a physical room and then a virtual replica of that room, using a suitable instrument provided by the field of environmental psychology. One of those instruments, the Semantic Environmental Description (Küller, 1972) has been used in the past to assess car interiors represented in different media including virtual reality (Karlsson, Aronsson, & Svensson, 2003). The virtual replica can also be presented in two levels of modelling skills: one amateur level and one professional level. One set of participants could perform a semantic environmental description for each of the three environments to identify differences between them. Another set of participants could be exposed to a VR experiment in each of the virtual environments (amateur and professional level) and differences in their response to the event could be observed.

Additionally, the visual realism of non-player characters can be studied further. The non-player characters included in the VR experiments presented here had a realistic look, but the effect of the uncanny valley (Mori & MacDorman, 2012) needs to be considered. It is possible, that if participants do not feel aversion towards the non-player characters, their interactions may be less aggressive. There are studies focusing the uncanny valley in VR (Schwind, 2018) and reaction to non-player characters of different kinds (Nelson, Mazumdar, Jamal, Chen, & Mousas, 2020). A study can be conducted in which different samples of participants are exposed to the same emergency in a virtual environment, with the style of the non-player characters varying per scenario: highly cartoonish, realistic, and hyper realistic characters. If participants feel aversion towards the non-player characters in their scenario, the effect of that aversion could be reflected in their behavior throughout the VR experience.

Addition of new technologies

New technologies can provide sensorial experiences modern VR equipment do not incorporate. Some studies have been conducted on the ability of VR to induce a stress response (Martens et al., 2019), but the vast majority of the studies are focused in the advantages VR offers to treat a variety of stress disorders, from social anxiety to post-traumatic stress disorder. Therefore, there is much to be studied in how to induce a mild stress response in participants exposed to dangerous scenarios. The use of external stressors should be considered. A more comprehensive study can be done to deepen previous findings on the use of radiative heat as a function of the height of the smoke layer (Blomander, 2020). New technologies, such as Cilia (Haptic Solutions Inc., 2020) allow the addition of smells to the virtual experience by a system of scents and fans. The smell of smoke could be added to a VR experiment including a fire, which could act as a first cue of the fire for the participant. However, it should be studied whether participants understand that the smell is part of the VR experience and not a physical-world incident taking place during the virtual environment. Moreover, some of the possible external stressors may have the downside of being compatible with only some of the commercially available VR platforms, or being in relatively immature states of development. These issues may burden the execution of the experiments with technical problems and incompatibilities that need to be solved.

The addition of external stressors means that the experiments would be moving from purely VR into augmented virtuality (AV). This migration may be necessary at least for some experiments that require the added stressors to make the virtual experiment more realistic. Studies may show if the added stressors work as expected and if the effect is large enough to justify the technical complexities they imply.

References

Abada, A., Abbrescia, M., AbdusSalam, S. S., Abdyukhanov, I., Abelleira Fernandez, J., Abramov, A., . . . Cook, C. T. A. (2019). FCC-hh: The Hadron Collider. *The European Physical Journal Special Topics, 228*(4), 755-1107. doi:https://doi.org/10.1140/epjst/e2019-900087-0

Andreatta, M., Neueder, D., Herzog, K., Genheimer, H., Schiele, M. A., Deckert, J., . . . Pauli, P. (2020). Generalization of conditioned contextual anxiety and the modulatory effects of anxiety sensitivity. *Neurotherapeutics*. doi:https://doi.org/10.1007/s13311-020-00831-8

Andrée, K., Nilsson, D., & Eriksson, J. (2016). Evacuation experiments in a virtual reality high-rise building: exit choice and waiting time for evacuation elevators. *Fire and Materials, 40*(4), 554-567. doi: https://doi.org/10.1002/fam.2310

Arias, S., Ronchi, E., Wahlqvist, J., Eriksson, J., & Nilsson, D. (2018). *ForensicVR: Investigating human behavior in fire with Virtual Reality* ((LUTVDG/TVBB; No. 3218)). Retrieved from Lund:

Arias, S., Ronchi, E., Wahlqvist, J., La Mendola, S., & Rios, O. (2019). Virtual Reality evacuation experiments on way-finding systems for the Future Circular Collider. *Fire Technology*. doi:https://doi.org/10.1007/s10694-019-00868-y

Best, R., & Demers, D. P. (1982). *Investigation report on the MGM Grand hotel fire*. Retrieved from Quincy, MA:

Boverkets byggregler (föreskrifter och allmänna råd), (2011).

Blascovich, J., Beall, A., Swinth, K., Hoyt, C., & Bailenson, J. (2002). Immersive Virtual Environment Technology as a Methodological Tool for Social Psychology. *Psychol. Inq, 13*.

Blomander, A. (2020). *Investigating the use of radiative heat panels to enhance perceived threat of fire in VR.* (Bachelor). Lund University, Lund, Sweden. (5608)

Blum, J., Rockstroh, C., & Göritz, A. S. (2019). Heart Rate Variability Biofeedback Based on Slow-Paced Breathing With Immersive Virtual Reality Nature Scenery. *Frontiers in Psychology, 10*(2172). doi:https://doi.org/10.3389/fpsyg.2019.02172

Bode, N. W. F., & Codling, E. A. J. F. T. (2018). Exploring Determinants of Pre-movement Delays in a Virtual Crowd Evacuation Experiment. doi:https://doi.org/10.1007/s10694-018-0744-9

Boletsis, C. (2017). The new era of virtual reality locomotion: A systematic literature review of techniques and a proposed typology. *Multimodal*

Technologies and Interaction, 1(4). doi:https://doi.org/10.3390/mti1040024

Boletsis, C., & Cedergren, J. E. (2019). VR Locomotion in the New Era of Virtual Reality: An Empirical Comparison of Prevalent Techniques. *Advances in Human-Computer Interaction, 2019*, 7420781. doi:https://doi.org/10.1155/2019/7420781

Bouchard, J. (1985). Fire in haunted castle kills eight. *NFPA Fire Journal, 79*(5), 11. Retrieved from https://www.nfpa.org/-/media/Files/News-and-Research/Publications-and-media/NFPA-Journal/May-June-2014/Haunted-Castle-Investigation-report.ashx

Browning, M. H. E. M., Mimnaugh, K. J., van Riper, C. J., Laurent, H. K., & LaValle, S. M. (2020). Can Simulated Nature Support Mental Health? Comparing Short, Single-Doses of 360-Degree Nature Videos in Virtual Reality With the Outdoors. *Frontiers in Psychology, 10*(2667). doi:https://doi.org/10.3389/fpsyg.2019.02667

Bryan, J. L. (1983). A review of the examination and analysis of the dynamics of human behavior in the fire at the MGM Grand Hotel, Clark County, Nevada as determined from a selected questionnaire population. *Fire Safety Journal, 5*(3-4), 233-240. doi:https://doi.org/10.1016/0379-7112(83)90021-8

Canter, D., Breaux, J., & Sime, J. (1980). *Domestic, Multiple Occupancy and Hospital Fires*. New York, NY: John Wiley & Sons.

Catal, C., Akbulut, A., Tunali, B., Ulug, E., & Ozturk, E. (2020). Evaluation of augmented reality technology for the design of an evacuation training game. *Virtual Reality, 24*(3), 359-368. doi:https://doi.org/10.1007/s10055-019-00410-z

Chang, C.-H., Pan, W.-W., Tseng, L.-Y., & Stoffregen, T. A. (2012). Postural activity and motion sickness during video game play in children and adults. *Experimental Brain Research, 217*(2), 299-309. doi:https://doi.org/10.1007/s00221-011-2993-4

Chardonnet, J.-R., Mirzaei, M. A., & Mérienne, F. (2017). Features of the Postural Sway Signal as Indicators to Estimate and Predict Visually Induced Motion Sickness in Virtual Reality. *International Journal of Human–Computer Interaction, 33*(10), 771-785. doi:https://doi.org/10.1080/10447318.2017.1286767

Chittaro, L., & Zangrando, N. (2010). The persuasive power of virtual reality: effects of simulated human distress on attitudes towards fire safety. *Persuasive Technology, 6137*(Persuasive 2010), 58-69. doi:https://doi.org/10.1007/978-3-642-13226-1_8

Coleman, R. (2017, 16 May 2017). Photorealistic VR is 20 years away. *PCGamesN*. Retrieved from https://www.pcgamesn.com/nvidia/nvidia-photorealistic-vr

Comeau, E., & Duval, R. F. (2000). *Fire investigation summary - Dance hall fire - Gothenburgh, Sweden*. Retrieved from Quincy, MA:

Cook, D., Dissanayake, D., & Kaur, K. (2019). Virtual Reality and Older Hands: Dexterity and accessibility in hand-held VR Control. *CHIuXiD'19*, 147-151. doi:https://doi.org/10.1145/3328243.3328262

Cosma, G., Ronchi, E., & Nilsson, D. (2016). Way-finding lighting systems for rail tunnel evacuation: A virtual reality experiment with Oculus Rift®. *Journal of Transportation Safety & Security*, 00-00. doi:https://doi.org/10.1080/19439962.2015.1046621

Craig, A. B., Sherman, W. R., & Will, J. D. (2009). Chapter 1 - Introduction to Virtual Reality. In A. B. Craig, W. R. Sherman, & J. D. Will (Eds.), *Developing Virtual Reality Applications* (pp. 1-32). Boston: Morgan Kaufmann.

Demetrescu, E., Ferdani, D., Dell'Unto, N., Leander Touati, A.-M., & Lindgren, S. (2016). Reconstructing the original splendour of the house of Caecilius Iucundus- A complete methodology for virtual archaeology aimed at digital exhibition. *Scientific Research and Information Technology, 6*(1), 51-66. doi:http://dx.doi.org/10.2423/i22394303v6n1p51

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *The Journal of Abnormal and Social Psychology, 51*(3), 629-636. doi:10.1037/h0046408

Donnely, C. (Writer) & C. Donelly (Director). (2006). Stardust. In K. Corrigan & N. Siderow (Producer). Ireland: Raidió Teilifís Éireann (RTÉ).

Duarte, E., Rebelo, F., Teles, J., & Wogalter, M. S. (2014). Behavioral compliance for dynamic versus static signs in an immersive virtual environment. *Applied Ergonomics, 45*(5), 1367-1375. doi:https://doi.org/10.1016/j.apergo.2013.10.004

Evans, L. (2019). A (brief) cultural history of VR. In *The re-emergence of Virtual Reality*. New York: Routledge.

Fernandes, A. S., & Feiner, S. K. (2016). *Combating VR sickness through subtle dynamic field of view modification.* Paper presented at the IEEE Symposium on 3D User Interfaces 2016, Greenville, SC, USA.

Fossataro, C., Tieri, G., Grollero, D., Bruno, V., & Garbarini, F. (2020). Hand blink reflex in virtual reality: The role of vision and proprioception in modulating defensive responses. *European Journal of Neuroscience, 51*(3), 937-951. doi:https://doi.org/10.1111/ejn.14601

Frantzich, H. (2001). *Experiments, Occupant Behaviour and Response Time – Results from Evacuation.* Paper presented at the Human Behaviour in Fire.

Fridolf, K., Nilsson, D., & Frantzich, H. (2011). Fire Evacuation in Underground Transportation Systems: A Review of Accidents and Empirical Research. *Fire Technology, 49*(2), 451-475. doi:https://doi.org/10.1007/s10694-011-0217-x

Fridolf, K., Ronchi, E., Nilsson, D., & Frantzich, H. (2013). Movement speed and exit choice in smoke-filled rail tunnels. *Fire Safety Journal, 59*, 8-21. doi:https://doi.org/10.1016/j.firesaf.2013.03.007

Gamberini, L., Chittaro, L., Spagnolli, A., & Carlesso, C. (2015). Psychological response to an emergency in virtual reality: Effects of victim ethnicity and emergency type on helping behavior and navigation. *Computers in Human Behavior, 48*, 104-113. doi:https://doi.org/10.1016/j.chb.2015.01.040

Gavgani, A. M., Walker, F. R., Hodgson, D. M., & Nalivaiko, E. (2018). A comparative study of cybersickness during exposure to virtual reality and "classic" motion sickness: are they different? *Journal of Applied Physiology, 125*(6), 1670-1680. doi:https://doi.org/10.1152/japplphysiol.00338.2018

Gibson, J. J. (1986). *The Ecological approach to visual perception*. Hillsdale (N.J.): Lawrence Erlbaum Associates.

Gromer, D., Reinke, M., Christner, I., & Pauli, P. (2019). Causal interactive links between presence and fear in Virtual Reality height exposure. *Frontiers in Psychology, 10*. doi:https://doi.org/10.3389/fpsyg.2019.00141

Grosshandler, W. L., Bryner, N. P., Madrzykowski, D., & Kuntz, K. (2005). *Report of the Technical Investigation of The Station Nightclub Fire* (NIST NCSTAR 2: Volume 1). Retrieved from Gaithersburg, MD (US):

Guinn, K. C. (1981). *The report of the governor's comission on fire safety codes*. Retrieved from Carson City, Nevada: http://fire.nv.gov/uploadedFiles/firenvgov/content/Resources/Commission%20on%20Fire%20Safety%20Codes%20-%20MGM%20Grand%20Final%20Report%2003111981.pdf

Guna, J., Geršak, G., Humar, I., Song, J., Drnovšek, J., & Pogačnik, M. (2019). Inlfuence of video content type on users' virtual reality sickness perception and physiological response. *Future Generation Computer Systems, 91*, 263-276. doi:https://doi.org/10.1016/j.future.2018.08.049

Gutierrez-Maldonado, J., Gutierrez-Martinez, O., Loreto, D., Peñaloza, C., & Nieto, R. (2010). Presence, involvement and efficacy of a virtual reality intervention on pain. *Studies in Health Technology and Informatics, 154*, 97-101.

Haptic Solutions Inc. (2020). Retrieved from https://hapticsol.com/

Heilig, M. L. (1962). United States of America Patent No.: U. S. P. Office.

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics, 6*(2), 65-70. Retrieved from http://www.jstor.org.ludwig.lub.lu.se/stable/4615733

Hoorn, J. F., Konijn, E. A., & van der Veer, G. C. (2003, February 2003). Virtual Reality: do not augment realism, augment relevance. *Upgrade - The European online magazine for the IT professional, IV(1),* 18-26. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.64.1652&rep=rep1&type=pdf#page=20

Ijsselsteijn, W., Nap, H. H., Poels, K., & De Kort, Y. (2007). *Digital Game Design for Elderly Users*.

Jenssen, G. D., Skjermo, J., Hoem, Å. S., Arnesen, P., Frantzich, H., & Nilsson, D. (2018). *Simulering av evakuering i tunnel* (2018:01000). Retrieved from http://hdl.handle.net/11250/2582549

Karlsson, B., Aronsson, N., & Svensson, K. (2003). Using semantic environment description as a tool to evaluate car interiors. *Ergonomics, 46*(13-14), 1408-1422.

Kinateder, M., Ronchi, E., Gromer, D., Müller, M., Jost, M., Nehfischer, M., . . . Pauli, P. (2014). Social influence on route choice in a virtual reality tunnel fire. *Transportation Research Part F: Traffic Psychology and Behaviour, 26*, 116-125. doi:https://doi.org/10.1016/j.trf.2014.06.003

Kinateder, M., Ronchi, E., Nilsson, D., Kobes, M., Müller, M., Pauli, P., & Mühlberger, A. (2014, 2014). *Virtual Reality for Fire Evacuation Research.* Paper presented at the Federated Conference on Computer Science and Information Systems.

Kinateder, M., & Warren, W. H. (2016). Social Influence on Evacuation Behavior in Real and Virtual Environments. *Frontiers in Robotics and AI, 3*. doi:https://doi.org/10.3389/frobt.2016.00043

Kinateder, M., Warren, W. H., & Schloss, K. B. (2019). What color are emergency exit signs? Egress behavior differs from verbal report. *Applied Ergonomics, 75*, 155-160. doi:https://doi.org/10.1016/j.apergo.2018.08.010

Kisker, J., Gruber, T., & Schöne, B. (2019). Behavioral realism and lifelike psychophysiological responses in virtual reality by the example of a height exposure. *Psychological Research*. doi:10.1007/s00426-019-01244-9

Kobes, M., Helsloot, I., de Vries, B., Post, J. G., Oberijé, N., & Groenewegen, K. (2010). Way finding during fire evacuation; an analysis of unannounced fire drills in a hotel at night. *Building and Environment, 45*(3), 537-548. doi:https://doi.org/10.1016/j.buildenv.2009.07.004

Kuligowski, E. (2016). Human behavior in fire. In M. J. Hurley, D. T. Gottuk, J. R. H. Jr., K. Harada, E. D. Kuligowski, M. Puchovsky, J. L. Torero, J. M. W. Jr., & C. J. Wieczorek (Eds.), *SFPE Handbook of Fire Protection Engineering*: Springer-Verlag New York.

Küller, R. (1972). *A semantic model for describing perceived environment*. Retrieved from National Swedish Institute for Building Research, Stockholm:

Lapinski, M. K., & Rimal, R. N. (2006). An Explication of Social Norms. *Communication Theory, 15*(2), 127-147. doi:https://doi.org/10.1111/j.1468-2885.2005.tb00329.x

Latané, B., & Darley, J. M. (1968). Group inhibition of bystanders intervention in emergencies. *Journal of Personality and Social Psychology, 10*, 215-221.

Latané, B., & Darley, J. M. (1970). *The unresponsive bystander: why doesn't he help?* New York: Appleton-Century Crofts.

Lawson, G., Shaw, E., Roper, T., Nilsson, T., Bajorunaite, L., & Batool, A. (2019). *Immersive Virtual Worlds: Multi-sensory Virtual Environments for Health and Safety Training*.

Liao, W., Kemloh Wagoum, A. U., & Bode, N. W. F. (2017). Route choice in pedestrians: determinants for initial choices and revising decisions. *Journal of The Royal Society Interface, 14*(127). doi:https://doi.org/10.1098/rsif.2016.0684

Månsson, J. (2018). *Using a virtual fire extinguisher as a tool for safety training.* (Master Master thesis). Lund University, Lund. Retrieved from http://lup.lub.lu.se/student-papers/record/8936710 (8936710)

Martens, M. A., Antley, A., Freeman, D., Slater, M., Harrison, P. J., & Tunbridge, E. M. (2019). It feels real: physiological responses to a stressful virtual reality environment and its impact on working memory. *Journal of Psychopharmacology, 33*(10), 1264-1273. doi:https://doi.org/10.1177/0269881119860156

MCV Staff. (2015). VR hardware needs to be '200 times more powerful' to achieve photorealism. *MCV/Develop*. Retrieved from https://www.mcvuk.com/development-news/vr-hardware-needs-to-be-200-times-more-powerful-to-achieve-photorealism/201850/

Milgram, P., & Colquhoun Jr., H. (1999). A taxonomy of real and virtual world display integration. In Y. Ohta & H. Tamura (Eds.), *Mixed reality - Merging real and virtual worlds* (pp. 5-30). Berlin: Springer-Verlag.

Mori, M., & MacDorman, K., Norri. (2012). The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine, 19(2),* 98-100.

Mossberg, A., Nilsson, D., & Andrée, K. (2020). Unannounced Evacuation Experiment in a High-Rise Hotel Building with Evacuation Elevators: A Study of Evacuation Behaviour Using Eye-Tracking. *Fire Technology*. doi:https://doi.org/10.1007/s10694-020-01046-1

Mossberg, A., Nilsson, D., & Wahlqvist, J. (2020). Evacuation elevators in an underground metro station: A Virtual Reality evacuation experiment. *Fire Safety Journal*, 103091. doi:https://doi.org/10.1016/j.firesaf.2020.103091

Moussaid, M., Kapadia, M., Thrash, T., Sumner, R. W., Gross, M., Helbing, D., & Hölscher, C. (2016). Crowd behaviour during high-stress evacuations in an immersive virtual environment. *Journal of the Royal Society Interface, 13*. doi:https://doi.org/10.1098/rsif.2016.0414

Munafo, J., Diedrick, M., & Stoffregen, T. A. (2017). The virtual reality head-mounted display Oculus Rift induces motion sickness and is sexist in its effects. *Experimental Brain Research, 235*, 889-901. doi:https://doi.org/10.1007/s00221-016-4846-7

Mütterlein, J. (2018, 03 Jan 2018). *The three pillars of Virtual Reality? Investigating the roles of immersion, presence and interactivity.* Paper presented at the 51st Hawaii International Conference on System Sciences, Hawaii.

Nelson, M. G., Mazumdar, A., Jamal, S., Chen, Y., & Mousas, C. (2020). *Walking in a Crowd Full of Virtual Characters: Effects of Virtual Character Appearance on Human Movement Behavior*, Cham.

Nguyen-Vo, T., Riecke, B. E., & Stuerzlinger, W. (2017, 18-19 March 2017). *Moving in a box: Improving spatial orientation in virtual reality using simulated reference frames.* Paper presented at the 2017 IEEE Symposium on 3D User Interfaces (3DUI).

Nilsson, D. (2009). *Exit choice in fire emergencies: influencing choice of exit with flashing lights*. Lund, Sweden: Dept. of Fire Safety Engineering and Systems Safety, Lund University.

Nilsson, D., Fridolf, K., & Frantzich, H. (2012). *Design of evacuation systems in underground transportation systems*. Paper presented at the Human Behavior in Fire international symposium 2012, Cambridge, United Kingdom. https://lup.lub.lu.se/record/3119341

Nilsson, D., & Johansson, A. (2009). Social influence during the initial phase of a fire evacuation—Analysis of evacuation experiments in a cinema theatre. *Fire Safety Journal, 44*(1), 71-79. doi:https://doi.org/10.1016/j.firesaf.2008.03.008

Proulx, G. (1993). A stress model for people facing a fire. *Journal of Environmental Psychology, 13*(2), 137-147.

Proulx, G. (2001). *Occupant behaviour and evacuation.* Paper presented at the 9th International Fire Protection Seminar, Munich.

Proulx, G., & Fahy, R. (1997). The Time Delay To Start Evacuation: Review Of Five Case Studies. *Fire Safety Science, 5*, 783-794. doi:https://doi.org/10.3801/IAFSS.FSS.5-783

Purser, D. A. (2016). Combustion toxicity. In M. J. Hurley, D. T. Gottuk, J. R. H. Jr., K. Harada, E. D. Kuligowski, M. Puchovsky, J. L. Torero, J. M. W. Jr., & C. J. Wieczorek (Eds.), *SFPE Handbook of Fire Protection Engineering*: Springer-Verlag New York.

Purser, D. A., & McAllister, J. L. (2016). Assessment of hazards to occupants from smoke, toxic gases and heat. In M. J. Hurley, D. T. Gottuk, J. R. H. Jr., K. Harada, E. D. Kuligowski, M. Puchovsky, J. L. Torero, J. M. W. Jr., & C. J. Wieczorek (Eds.), *SFPE Handbook of Fire Protection Engineering* (5th edition ed.): Springer-Verlag New York.

Rebenitsch, L., & Owen, C. (2016). Review on cybersickness in applications and visual displays. *Virtual Reality, 20*(2), 101-125. doi:https://doi.org/10.1007/s10055-016-0285-9

Ronchi, E., Kinateder, M., Müller, M., Jost, M., Nehfischer, M., Pauli, P., & Mühlberger, A. (2015). Evacuation travel paths in virtual reality experiments for tunnel safety analysis. *Fire Safety Journal, 71*, 257-267. doi:https://doi.org/10.1016/j.firesaf.2014.11.005

Ronchi, E., & Nilsson, D. (2015). *A Virtual Reality experiment on the design of flashing lights at emergency exit portals for road tunnel evacuations* (3180). Retrieved from Lund [Sweden]:

Ronchi, E., Nilsson, D., Kojić, S., Eriksson, J., Lovreglio, R., Modig, H., & Walter, A. L. (2015). A Virtual Reality Experiment on Flashing Lights at Emergency Exit Portals for Road Tunnel Evacuation. *Fire Technology*. doi:https://doi.org/10.1007/s10694-015-0462-5

Runefors, M., Johansson, N., & Van Hees, P. (2016). How could the fire fatalities have been prevented? An analysis of 144 cases during 2011-2014 in Sweden. *Journal of Fire Sciences, 34*(6), 515-527. doi:https://doi.org/10.1177/0734904116667962

Saredakis, D., Szpak, A., Birckhead, B., Keage, H. A. D., Rizzo, A., & Loetscher, T. (2020). Factors Associated With Virtual Reality Sickness in Head-Mounted Displays: A Systematic Review and Meta-Analysis. *Frontiers in human neuroscience, 14*, 96-96. doi:https://doi.org/10.3389/fnhum.2020.00096

Saunders, J., Bodempudi, S. T., Sharma, S., Scribner, D., Grynovicki, J., & Grazaitis, P. (2018). *Building evacuation using Microsoft HoloLens*. Paper presented at the SEDE 2018, New Orleans, Louisiana.

Schutte, N. S., & Stilinović, E. J. (2017). Facilitating empathy through virtual reality. *Motivation and Emotion, 41*, 708-712. doi:https://doi.org/10.1007/s11031-017-9641-7

Schwind, V. (2018). *Implications of the uncanny valley of avatars and virtual characters for human-computer interaction.* (Doctoral monography). University of Stuttgart, Stuttgart. Retrieved from https://elib.uni-stuttgart.de/handle/11682/9953

Sharma, S., Devreaux, P., Sree, S., Scribner, D., Grynovicki, J., & Grazaitis, P. (2019). Artificial intelligence agents for crowd simulation in an immersive environment for emergency response. *Electronic Imaging, 2019*(2), 176-171-176-178. doi:https://doi.org/10.2352/ISSN.2470-1173.2019.2.ERVR-176

Shaw, E., Roper, T., Nilsson, T., Lawson, G., Cobb, S. V. G., & Miller, D. (2019). *The Heat is On: Exploring User Behaviour in a Multisensory Virtual Environment for Fire Evacuation*. Paper presented at the Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, Scotland Uk. https://doi.org/10.1145/3290605.3300856

Shearer, E., & Gottfried, J. (2017). Americans' online news use is closing in on TV news use [online article]. *News use across social media platforms 2017*. Retrieved from http://www.pewresearch.org/fact-tank/2017/09/07/americans-online-news-use-vs-tv-news-use/

Shields, T. J., & Proulx, G. (2000). *The science of human behaviour: past research endeavours, current developments and fashioning a research agenda.* Paper presented at the Fire Safety Science - 6th International Symposium.

Sime, J. D. (1984). *Escape behaviour in fires: "panic" or affliation?* (Doctor of Philosophy). University of Surrey, Retrieved from http://epubs.surrey.ac.uk/id/eprint/848024

Sime, J. D. (1985). Movement toward the Familiar: Person and Place Affiliation in a Fire Entrapment Setting. *Environment and Behavior, 17*(6), 697-724. doi:https://doi.org/10.1177/0013916585176003

Slater, M. (2003). A note on presence terminology. In (Vol. 3): Presence Connect.

Slater, M., Usoh, M., & Steed, A. (1995). Taking steps: the influence of a walking technique on presence in virtual reality. *ACM Transactions on Computer-Human Interaction, 2*(3), 201-219.

Statens medieråd. (2019). *Unger & medier 2019* (ISSN 2001-6840). Retrieved from Stockholm: https://statensmedierad.se/download/18.126747f416d00e1ba946903a/1568041620554/Ungar%20och%20medier%202019%20tillganglighetsanpassad.pdf

Steuer, J. (1992). Defining Virtual Reality: Dimensions Determining Telepresence. *Journal of Communication, 42*(4), 73-93. doi:https://doi.org/10.1111/j.1460-2466.1992.tb00812.x

Stevenson Won, A., Bailenson, J., Lee, J., & Lanier, J. (2015). Homunlar flexibility in virtual reality. *Journal of Computer-Mediated Communication, 20*(3), 241-259. doi:https://doi.org/10.1111/jcc4.12107

Sutherland, I. E. (1968). *A head-mounted three dimensional display*. Paper presented at the Proceedings of the December 9-11, 1968, fall joint computer conference, part I, San Francisco, California. https://doi.org/10.1145/1476589.1476686

Swartz, J. A. (1979). Human behavior in the Beverly Hills fire. *NFPA Fire Journal, 79*(3), 73-74. Retrieved from https://www.nfpa.org/-/media/Files/News-and-Research/Resources/Fire-Investigations/FISOUTHGATE.ashx?as=1&iar=1&hash=6F396F91ED67A36CD0D78C58559ACDE6

Tancogne-Dejean, M., & Laclémence, P. (2016). Fire risk perception and building evacuation by vulnerable persons: Points of view of laypersons, fire victims and experts. *Fire Safety Journal, 80*, 9-19. doi:https://doi.org/10.1016/j.firesaf.2015.11.009

Tang, C.-H., Wu, W.-T., & Lin, C.-Y. (2009). Using virtual reality to determine how emergency signs facilitate way-finding. *Applied Ergonomics, 40*(4), 722-730. doi:https://doi.org/10.1016/j.apergo.2008.06.009

Teague, P. E., & Farr, R. R. (2009). Supplement 1 Case Histories: Fires Influencing the Life Safety Code. In N. F. P. Association (Ed.), *Life Safety Code Handbook* (pp. 1147-1159). Quincy, MA: National Fire Protection Association.

Tong, D., & Canter, D. (1985). The decision to evacuate: a study of the motivations which contribute to evacuation in the event of fire. *Fire Safety Journal, 9*(3), 257-265. doi:https://doi.org/10.1016/0379-7112(85)90036-0

Tribunal of inquiry of the fire at the Stardust, & Keane, R. (1982). *Report of the Tribunal of inquiry of the fire at the Stardust, Artane, Dublin on the 14th February, 1981*. (PL 853). Dublin: The Stationary Office

Troncoso, J., Nilsson, D., & Ronchi, E. (2015, 2015). *Response to Emergency Way-finding Systems by People from Different Cultures.* Paper presented at the Human Behaviour in Fire Symposium.

Wahlqvist, J., & van Hees, P. (2018). *Visualization of fires in VR* (3222). Retrieved from Lund: https://www.brandforsk.se/wp-content/uploads/2020/02/Brandforsk_Rapport_Simulering_Virtual_Reality.pdf

Wann, J. P., Rushton, S., & Mon-Williams, M. (1995). Natural problems for stereoscopic depth perception in virtual environments. *Vision Research, 35*(19), 2731-2736. doi:https://doi.org/10.1016/0042-6989(95)00018-U

Weech, S., Varghese, J. P., & Barnett-Cowan, M. (2018). Estimating the sensorimotor components of cybersickness. *Journal of Neurophysiology, 120*(5), 2201-2217. doi:https://doi.org/10.1152/jn.00477.2018

Wetterberg, N., Ronchi, E., & Wahlqvist, J. (2020). Individual Driving Behaviour in Wildfire Smoke. *Fire Technology*. doi:https://doi.org/10.1007/s10694-020-01026-5

Wright, J., & van Waveren, J. M. P. (Producer). (2014, July 2020). Getting started with the Oculus Mobile SDK. *Samsung Developer Conference 2014*. Retrieved from https://www.youtube.com/watch?v=PiLOX06sobc

Yildirim, C. (2020). Don’t make me sick: investigating the incidence of cybersickness in commercial virtual reality headsets. *Virtual Reality, 24*(2), 231-239. doi:https://doi.org/10.1007/s10055-019-00401-0

Zibrek, K., Martin, S., & McDonnell, R. (2019). Is photorealism important for perception of expressive virtual humans in virtual reality? *ACM Transactions on Applied Perception, 16*(3). doi:https://doi.org/10.1145/3349609

Zihl, J. (2006). Zerebrale Blindheit und Gesichtsfeldausfälle. In H.-O. Karnath & P. Thier (Eds.), *Neuropsychologie* (pp. 88-96). Berlin, Heidelberg: Springer Berlin Heidelberg.