Fire Technology, 58, 817–849, 2022

Manufactured in The United States
https://doi.org/10.1007/s10694-021-01172-4

# A Study on Evacuation Behavior in Physical and Virtual Reality Experiments

*Silvia Arias* *, *Division of Fire Safety Engineering, Faculty of Engineering, Lund University, Lund, Sweden*
*Axel Mossberg, Division of Fire Safety Engineering, Faculty of Engineering, Lund University, Lund, Sweden and Bengt Dahlgren Fire Research, Krokslätts fabriker 52, 431 37 Mölndal, Sweden*
*Daniel Nilsson, Civil and Natural Resources Engineering, University of Canterbury, Christchurch, New Zealand*
*Jonathan Wahlqvist, Division of Fire Safety Engineering, Faculty of Engineering, Lund University, Lund, Sweden*

**Received:** 29 September 2020/**Accepted:** 26 August 2021/**Published online:** 17 September 2021

***Abstract.*** Comparing results obtained in Virtual Reality to those obtained in physical experiments is key for validation of Virtual Reality as a research method in the field of Human Behavior in Fire. A series of experiments based on similar evacuation scenarios in a high-rise building with evacuation elevators was conducted. The experiments consisted of a physical experiment in a building, and two Virtual Reality experiments in a virtual representation of the same building: one using a Cave Automatic Virtual Environment (CAVE), and one using a head-mounted display (HMD). The data obtained in the HMD experiment is compared to data obtained in the CAVE and physical experiment. The three datasets were compared in terms of pre-evacuation time, noticing escape routes, walking paths, exit choice, waiting times for the elevators and eye-tracking data related to emergency signage. The HMD experiment was able to reproduce the data obtained in the physical experiment in terms of pre-evacuation time and exit choice, but there were large differences with the results from the CAVE experiment. Possible factors affecting the data produced using Virtual Reality are identified, such as spatial orientation and movement in the virtual environment.

**Keywords:** Evacuation, Elevators, Virtual Reality, Fire safety, Eye-tracking

## 1. Introduction

Virtual Reality (VR) has been used for data collection in the field of Human Behavior in Fire for many years. Several studies have been conducted on way-finding [8, 11], behavior during emergencies [5, 12], and system design [11, 14, 15, 19]. The extrapolation of results from VR to the real-world evacuation, however,

*Correspondence should be addressed to: Silvia Arias, E-mail: silvia.arias@brand.lth.se

needs special attention. Studies are being performed to determine how closely VR can replicate data obtained using other research methods [2–4, 10, 16], but a broad body of research is needed to validate the VR experiment method.

No single experiment can cover all possible aspects of the VR experiment method to be validated. Some validation studies may compare the VR data to that from real fire emergencies [2], which are an excellent source for behavioral data but lack experimental control [9]. As an alternative, a physical experiment (either laboratory or field experiment) can be replicated in VR [10, 16]. In this case, the high level of experimental control of the physical experiment matches that of VR. This match allows for more precise comparisons between the datasets, but the ecological validity of the data may be lower in VR than in physical experiments [9]. These limitations are intrinsic to the source of data and a recurrent theme in validation of any method. Nevertheless, for VR to be validated, it needs to be able to produce data at least as reliable as other research methods. Therefore, its comparison with a physical experiment is relevant to the validation process.

Some real-world experiments are more suitable for comparison with VR than others, as VR has limitations in the data that it can collect. For example, walking speeds are usually measured in different physical experiments. However, they can be difficult to accurately represent in VR unless the participant can move as naturally as they would in a physical environment. This natural movement is easy to implement in small virtual environments using a system of cables or a wireless solution. Nevertheless, the larger the environment, the more challenging it becomes to capture natural movement. There may be limitations in the area of coverage of the VR equipment used, or even the availability of a large enough and safe experimental room. High occupant density scenarios can also be hard to replicate in VR, as physical contact with virtual building occupants can be hard to represent accurately in the experimental room. Moreover, being the VR experience heavily reliant on visual inputs, scenarios with very low visibility conditions (such as dark environments due to power outage or heavy smoke conditions) may offer too little input for the participant to find their way out or make any decisions on how to act. Therefore, the selection of a physical experiment for the VR method to be validated against needs to take into consideration the limitations of the VR method. The identification of a suitable physical experiment was the first step to launch this validation study.

As mentioned before, other studies attempting to reproduce physical experiments in VR have been conducted by different groups of researchers with overall positive results [10, 16]. Despite their level of success, it should be noted that any VR experiment aiming to reproduce a given physical experiment cannot be considered a source of data but a proof of concept. This is because the application of VR for research in fire evacuation is still under development, and much is yet to be understood in terms of the differences between VR and reality. VR is by no means reality, and the studies trying to reproduce reality in VR mean to identify differences, propose solutions to mitigate their effect, or single out the limitations of the VR method [3, 4]. These replication studies are needed for the implementa-

tion of VR experiments as a research method for Human Behavior in Fire research to be validated in the future.

The present paper is a study on replicability of a physical experiment in VR. Therefore, it should not be considered a source of behavioral data to be included in a model. As such, the data presented here should not be considered representative of the evacuation behavior of hotel occupants. This is especially valid since the physical experiment conducted by Mossberg, Nilsson, and Andrée [14] also does not consider their sample of participants representative of the expected occupancy type in a hotel. The differences they list include the fact that participants signed up for the experiment and came to the hotel under the guise of a study on interior design, had no luggage with them, were awake and had only been in the hotel for a short while before the alarm went off. These conditions of the sample in the physical experiment were roughly the same for participants in the VR experiments.

Moreover, the data obtained from VR experiments does not need to exactly replicate the data from a physical experiment, since not even the repetition of said physical experiment is able to replicate the data from the first iteration. There are many factors playing a role in the decision-making of individuals in a fire event (e.g., background, previous experience, risk perception, personality traits, etc.). Therefore, it is not necessary for the data obtained from VR experiments to show no statistically significant difference. Differences are expected, but dramatically large differences may indicate a problem.

The objective of this paper is to compare the results obtained from three experiments (one physical and two VR experiments) based in similar scenarios in order to assess the suitability of VR to replicate real-world behavior. The data produced in the two VR experiments will also be compared with each other, to assess differences in the data produced using different types of VR equipment. In addition, the perceived realism of the most recent VR experiment from the participants' point of view will be evaluated. This evaluation aims at exploring how realistic the VR experience was for them, which may indicate whether the VR experiment succeeded at simulating a credible emergency scenario.

## 2. Method

Three experiments will be compared with each other to assess how similar the data obtained in VR is to that of the physical experiment. The experiments were based on an existing high-rise building located in Stockholm. The building has the special feature of relying on the use of elevators as means of egress in an evacuation. The experiments had the same building layout and similar scenarios. They took place in different years. The first one, run by Andrée, Nilsson, and Eriksson [1], was performed in VR. This experiment is referred to here as *the CAVE experiment*, because it used a Cave Automatic Virtual Environment (CAVE). The second experiment was conducted by Mossberg, Nilsson, and Andrée [14] is referred to here as *the physical experiment*, because it took place in the existing building. The third experiment is referred to as *the HMD experiment*, because it was per-

formed in VR using a Head Mounted Display (HMD), and it has not been published before this paper. Therefore, especial detail will be given to the description of the HMD experiment. Summarized descriptions of the other two experiments will be provided, along with the corresponding references for further detail.

### 2.1. The Real Building

The existing building is a 35-story mixed-occupancy building, which includes hotel and office floors. It was designed to rely on a single emergency staircase and a set of evacuation elevators. The general layout of a hotel floor in this building is shown on Fig. 1. All six elevators are configured to operate during an evacuation, and the elevator lobby is designed as a safe area. The doors leading to the elevator lobby are connected to the fire alarm, and they self-close once the alarm is triggered to prevent the spread of smoke into the elevator lobby. One emergency exit sign (indicated in Fig. 1 by a cross) is attached to the outside face of one of the door leaves, so that building occupants can identify the elevator lobby also when the doors are closed. In addition, a voice alarm announces that it is safe to use the elevators for evacuation in this building. The voice alarm consists of a sequence of beeps, followed by a male voice giving a message in Swedish, a second sequence of beeps, and then the same voice giving the same message in English. The English message says.

> "Attention please. Attention please. There is a fire in the building. Evacuate immediately through nearest exit. The lifts can be used as an emergency exit in this building".

The entirety of the alarm is played in a loop until the emergency is taken care of. The length of one cycle of the alarm is 43 s, equally distributed between the two languages. All features described here were included in the three experiments.

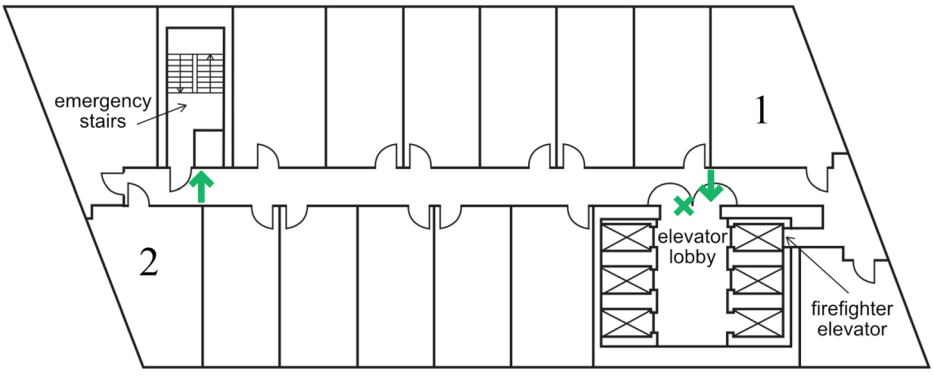


**Figure 1. Layout of floor 16 in the hotel section of the building, indicating the location of the emergency stairs, the elevator lobby and the firefighter elevator. The arrows and the cross indicate the location of emergency exit signage and the direction they pointed at.**

### *2.2. Shared Methodology*

The three experiments aimed to provide data on the evacuation behavior of one participant at a time when an alarm went off in the high-rise hotel building. Invariably, the participant started the experiment in the lobby on the ground floor, where they were given instructions to go to their assigned room on floor 16. To do so, the participant used the elevators and went to the assigned room by themselves. Participants were not told that the real purpose of the experiment was to observe their evacuation behavior, to avoid bias in their behavior. Instead, they were given a bogus task to perform in their assigned room, so that they could be caught off guard by the alarm. The alarm was manually triggered when the participant was in the room. The alarm played in a loop, non-stop, until the end of the experiment. The participant then had to decide on their own whether to evacuate, and whether to use the elevators or the stairs for that. Their choices were recorded. Once the experiment ended, the participant was asked to fill in a questionnaire, and the debriefing session followed afterwards.

### *2.3. CAVE Experiment*

The first VR experiment, the CAVE experiment, aimed to investigate the exit choice and the waiting time in a building like this high-rise hotel building [1]. The CAVE used consists of a room-size set of screens (three as walls, one on the floor), in which the participant stands. The virtual environment is projected on those screens and the participant navigates it using a hand controller. Figure 2 illustrates a CAVE system with four screens and projectors, as used by Andrée et al. [1].

A total of 72 participants were distributed in two scenarios in this experiment: a baseline scenario, and one in which the emergency signage pointing at the elevator was enhanced by the addition of flashing lights. In both scenarios, participants

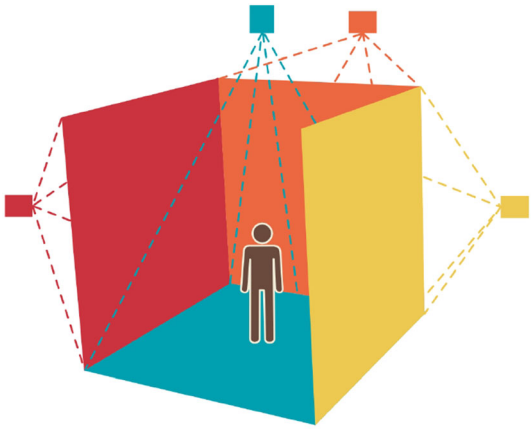

**Figure 2. Representation of the CAVE system used in the CAVE experiment and person is standing in it.**

were assigned to room 1 (see Fig. 1). The experiment was designed so that participants choosing the elevators as means of egress would have to wait up to 20 min without an elevator stopping by the floor they were in. This feature allowed for waiting times to be collected. Those participants could decide at any point not to wait any longer and go to the stairs.

The results provided insight on whether participants would use the elevators or the stairs, and how long they were willing to wait for the elevators in the elevator lobby. They also showed two possible sources of bias. Firstly, being located in room 1 (see Fig. 1) in the vicinity of the elevator lobby, and having the participant been given no chance to explore the rest of the hallway before the alarm went off, participants may have been biased to choose the elevator (being close and relatively familiar). The second possible source of bias was the difficulty participants experienced navigating in the virtual environment. In the CAVE they stood still and navigated the virtual environment using a hand-controller. The navigation system proved to be somewhat difficult, especially when turning around, which may have led participants to not to change the path they chose at first.

### *2.4. Physical Experiment*

A few years later, the physical experiment aimed to reproduce the CAVE experiment in the real building. A total of 67 participants were distributed in three scenarios. In addition to the two scenarios used in the CAVE experiment, the physical experiment had a third scenario, in which participants were assigned to room 2 (see Fig. 1. This scenario would show whether the vicinity of the elevator affects the decision of using it. Other than that, participants were equipped with an eye-tracking device (Tobii Pro Glasses 2 with output analysis in the software Tobii Pro Lab), to collect data on the features catching their attention during the evacuation (e.g., noticing emergency signage). The fixation velocity threshold for the eye-tracking glasses was set to 100 degrees/second and a minimum fixation duration of 60 ms. To justify the use of this equipment without revealing the real purpose of the experiment, participants were told that the experiment was about the interior design of the assigned room.

To reduce the impact of the experiment in business operation, the voice alarm was only played on floor 16, which hosted no guests other than the participant. The experiment ended when the participant entered the staircase or when they pressed the call button for the elevators. No data on waiting time was recorded to prevent a confusion ensuing between the participant and other guests using the elevators as usual, unaware of an experiment taking place at the time, or of an alarm going off at all. The results provided data on pre-evacuation time, exit choice, walking paths and eye movements during the evacuation.

### *2.5. HMD Experiment*

The HMD experiment aimed at collecting the same data as the previous two. In order to collect data on eye movement as the physical experiment did, an eye-tracking device was added to the HMD. Objects of interest were defined as those

the participant could look at to get information relevant to the evacuation, as they were also defined in the physical experiment. Examples of objects of interest were the emergency exit signage and the evacuation plans (in the hotel room and in the hallway). The HMD experiment was designed to collect waiting time for the elevators, as the CAVE experiment did.

*2.5.1. Participants* Participants were recruited via announcement on a specialized website for recruitment of participants regularly used by researchers at Swedish universities. Table 1 presents demographical information about the participants in the HMD experiment. No participant was a student at a Fire Engineering program taught at Lund University, neither at bachelor nor master level. As a compensation for their involvement, each participant received a single movie ticket, worth around 100 SEK.

*2.5.2. Equipment* The equipment consisted of a commercially available VR head-mounted display (HMD) with its hand controllers (HTC Vive™, dual AMOLED 3.6″ diagonal screen, 1080 × 1200 pixels per eye, 90 Hz refresh rate and 110° of field of view, and 6 degrees of freedom of movement), and an eye-tracking add-on specifically made for HTC Vive™ (Pupil labs HTC Vive Binocular add-on, 192 × 192 pixels per eye, binocular 200 Hz tracking frequency, ∼ 1.0° gaze accuracy, ∼ 0.08° gaze precision, 5.7 ms latency).

The room configuration of the VR equipment was *standing only*, which means the participants stood in the same position in the experimental room during the whole experiment. The hand controllers were used for navigation in the virtual environment and for interacting with the objects in it. A demonstration of the use of the hand controllers was conducted before the participants were fitted with the equipment.

A high-end gaming computer was used to meet the performance requirements of the HMD. The computer included an Intel i7 7700 k CPU, and Nvidia GeForce GTX 1080 8 GB GPU and 32 GB of RAM. The computer could keep a locked framerate of 90 frames per second throughout the experiment, in order to avoid inconsistencies between the participants' movements and the rendered ima-

**Table 1**
**Demographical Description of the Participants, Showing the Composition of the Samples in each Scenario**

| | Gender | | | Age | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| *Sample* | *Female* | *Male* | *Other* | *Min* | *Max* | *Average* | *Std dev* | *Mode* | *Median* |
| Scenario 1 | 16 | 14 | 1 | 18 | 40 | 25.35 | 5.10 | 22 | 24 |
| Scenario 2 | 17 | 14 | 0 | 19 | 58 | 25.32 | 7.65 | 23 | 23 |
| Total | 33 | 28 | 1 | 18 | 58 | 25.34 | 6.45 | 23 | 23 |

ges. The virtual environment was generated using SketchUp™ version 2017, and the game engine used was Unity 3D™ (version 2019.1.0f2).

With the data produced by the eye-trackers, it was possible to determine the number of times the participant looked at an object of interest, and for how long they looked at it each time. The precision of this measurement was 0.01111 s. The data showed whether they looked at the object of interest before or after the activation of the alarm. Data was collected also on how much time the participant spent in the hotel room after the alarm was triggered, the time spent waiting in the elevator lobby, and the total time since the alarm went off.

*2.5.3. Scenarios* Two scenarios were included in this experiment, namely Scenario 1 and Scenario2. In Scenario 1, participants were assigned to room 1, and in Scenario 2 they were assigned to room 2 (see Fig. 1). These scenarios matched the location of those in the physical experiment. The layout of the lobby and the rooms on the 16$^{th}$ floor was replicated in VR in a 1:1 scale. The location of emergency signage was also replicated. The lighting conditions in the virtual environment were similar to those in the real building. Some parts of the real building were easily replicated in VR, such as the hallway on floor 16, and elevator and the elevator lobby. Other parts, such as the lobby and the hotel rooms, were not identical in appearance but presented a general resemblance. Figure 3 shows a view from the virtual environment alongside a photo of the real building.

There were two minor differences in the layout of the virtual environment compared to the real building. In the real building, one of the elevators can be accessed directly from the hallway, without entering the elevator lobby. This access was eliminated in the virtual environment for simplicity, as it was also not included in the CAVE experiment. The other difference was a second (internal) door on the staircase, which in the real building constitutes an airlock. The internal door was

**Figure 3. View of the hallway on floor 16 used for the physical experiment (left hand side) and its virtual representation used in the VR experiment. The view is seen from room 1 and along the corridor having the elevator lobby to the left.**

removed, leaving only the one on the hallway, because of observed difficulties on participants operating swinging doors in VR. By removing this door, participants only needed to open the one on the hallway to see they were entering a staircase. These differences are shown in Fig. 4.

*2.5.4. Procedure* Participants were recruited under the pretense of an experiment about assessing realism of a ''highly realistic VR hotel''. Each participant arrived at the time and date they chose for themselves, one at a time. The participant was asked to sign the informed consent form upon arrival. Then, a demonstration of the use of the VR equipment was given. The participant was also made aware of the eye-tracking device on the HMD, and it was explained to them that a calibration of the device would be conducted once the VR scenario was launched. They were asked if they had any questions, and once they were content with the information they received, they were asked to stand in the designated location to start the experiment. There, they fitted themselves with the equipment and one of the two scenarios was launched.

The participant could immediately see the virtual environment they were in, which was the lobby of a hotel. They were then given instructions on how to do the calibration of the eye-tracking device. The calibration took approximately 30 s, and a message on the computer screen indicated if it was successful. Once the calibration was completed, the participant was given instructions about what to do next. They were told they would first receive a full set of instructions and then they would be on their own. The instructions indicated the participant needed to follow the arrows they could see on the floor to get to the elevator lobby. Once there, they needed go to floor 16, and were given the room number to go to depending on which scenario the participant was in. They were told they were supposed to go to the room, which was unlocked, and once there they were sup-

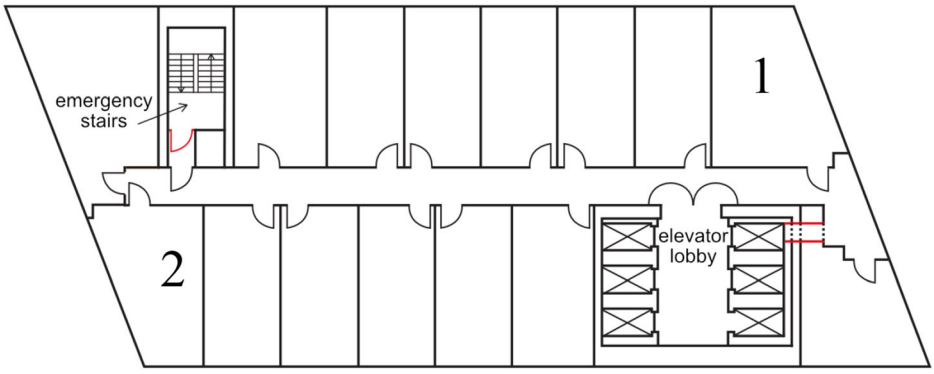


**Figure 4. Layout of the real building indicating in red the features that were not included in the virtual environment: an access to one of the elevators directly from the hallway (see dead end on the right-hand side of the figure), and an internal door (airlock) in the staircase.**

posed to pay attention to the room. They were told the virtual hotel room was as realistic as a real one, and that they were supposed to assess its realism. They were also encouraged to try and use different objects in the room as they would do in reality. The participant was reminded they could interrupt the experiment at any time, for any possible reason. They were also told that the computer screen was mirroring the images they were seeing, and that the screen was being recorded. Then they were asked again if they had questions, and if not, they were free to start.

The participant started by following the arrows to the elevator lobby on the ground floor, where one out of six elevators was on hold with the doors open. Once the participant got in the elevator and pressed the button to floor 16, the elevator doors closed. A small screen in the elevator showed the number of the floors the elevator was passing by. Once it reached the 16th floor, the doors opened. The participant arrived to an elevator lobby, also fitted with six elevators (see Fig. 1). At this point, they were verbally reminded of the number of the room they were supposed to go to. A sign on the wall across the hallway indicated whether the room they were looking for was towards their left or their right. The participant then walked down the hallway and reached their room. They accessed it by pulling the door handle with a hand controller, similar to opening a regular door. The participant was given time to look around the room and try some of the objects in it. The fire alarm was launched manually and without notice while the participant seemed to be focused in exploring the room.

The participant then made their decision to evacuate based on the message given by the voice alarm. An evacuation plan was mounted on the inside of the hotel room door, in which they could see the location of the stairs and the elevators. Once in the hallway, they could see signage indicating the emergency routes, as shown in Fig. 1. If the participant went for the stairs, the experiment ended once they reached the first flight of stairs. If they went for the elevators, they had to press the call buttons for the elevators, and wait. Unbeknown to the participant, the elevators were programmed to move between the floors but never stop on floor 16. The participant had two options: keep waiting for an elevator, or take the stairs. If they chose the first, they would be left waiting for up to 5 min before ending the experiment. If they chose the stairs, it ended when they reached the first flight.

The participant was asked then to remove the equipment, and to fill in an online questionnaire. In the meantime, the eye-tracking data and the video were stored safely in the computer. Upon completion of the questionnaire, the debriefing took place, allowing the participant to ask questions. Finally, the participant was thanked for their time with a movie ticket, and were walked to the exit.

*2.5.5. Analysis* Participants could choose to use the stairs or the elevators as their means of egress. This exit choice was classified as ''stairs'' if they entered the emergency staircase; and ''elevator'' if they pressed an elevator call-button. The elevator classification was furthermore divided on whether they waited for the elevators for five minutes (''elevator only'') or they decided not to wait any longer and switched to the stairs instead (''switched to stairs''). To compare the results of

the HMD experiment to those of the physical experiment, the ''elevator only'' and the ''switched to stairs'' classes were combined.

The videos made of the computer screen were watched in order to piece together the walking paths of the participants. Milestones were defined, such as ''room'', ''dead-end'', ''passed by elevator lobby'', ''(elevator) lobby'', ''stairs'', to establish sequences of the trajectory. The time it took for each participant to reach a milestone was not recorded, but the number of times they went to one was (e.g. entering the elevator lobby, going to the dead-end, entering the elevator lobby again).

The data obtained from the eye-tracking equipment in the HMD experiment only referred to how long the gaze intersected the object of interest. The precision of 0.01111 s allowed for defining a gaze fixation as a period equal or longer than 0.06 s, which was the length of the gaze fixation adopted in the physical experiment. Moreover, the data collected contained information about how many times the participant fixated the gaze on the object of interest and for how long each time. Therefore, gazing at an object of interest for less than 0.06 s was not counted as a gaze fixation.

## 3. Summary Comparison of the Experiments

To summarize the information presented in the previous sections, the small variations between the experiments are highlighted here. Table 2 presents an overview of the participants in the three samples. Table 3 presents a summary of the experiments, indicating also the data collected in each, to be presented in the results section.

**Table 2**
**Overview of the Demographics of the Participants in each Experiment in Terms of Age, Gender and Recruitment**

| | Gender | | | Age | | | |
|---|---|---|---|---|---|---|---|
| Experiment | *Female* | *Male* | *Other* | *Min* | *Max* | *Mean* | Description |
| CAVE | 29 | 43 | – | 18 | 69 | 26.5 | Lund University students and staff recruited through emails and in person – no fire safety students or staff |
| Physical | 37 | 30 | – | 20 | 71 | 33.2 | general public recruited online |
| HMD | 33 | 28 | 1 | 18 | 58 | 25.3 | general public recruited online – no fire safety students or staff |

**Table 3**
**Summary of Similarities and Differences Between the Three Experiments. *The Flashing Lights Scenario was not Included in the HMD Experiment, and Therefore was Excluded from the Data Analysis**

| Experiment | CAVE | Physical | HMD |
|---|---|---|---|
| Method | VR | Real-world | VR |
| Publication | Andrée et al. [1] | Mossberg, Nilsson, and Andrée [14] | Present paper |
| Participants | | | |
| *Scenario 1 (room 1)* | 52 | 22 | 31 |
| *Scenario 2 (room 2)* | 0 | 23 | 31 |
| *Flashing lights** | 20 | 22 | 0 |
| Equipment | CAVE | Eye-tracking device | HMD with eye-trackers |
| Data collected | | | |
| *Exit choice* | Yes | Yes | Yes |
| *Waiting time* | Yes | No | Yes |
| *Pre-evacuation time* | No | Yes | Yes |
| *Walking path* | No | Yes | Yes |
| *Eye-movement* | No | Yes | Yes |

## 4. Results

In this section, results obtained from the HMD experiment will be presented and compared to those from both the CAVE experiment and the physical experiment. The results presented in this paper comprise the following aspects:

(1) Pre-evacuation time
(2) Noticing available escape routes
(3) Walking paths
(4) Exit choice
(5) Waiting time
(6) Eye-tracking data
(7) VR-related results

As a reminder, Table 4 presents an overview of the results produced by each experiment to be compared with each other. The present paper will not compare any data collected from the flashing light scenarios used in the CAVE and the physical experiment. The CAVE experiment did not collect data about pre-evacuation time, participants being aware of the available escape routes, or walking paths. No eye-tracking device was used in that experiment either.

**Table 4**
**Overview of the Results from each Experiment to be Compared with each Other**

| Results | Physical | HMD | CAVE |
|---|---|---|---|
| *Pre-evacuation time* | × | × | |
| *Noticing available escape routes* | × | × | |
| *Walking paths* | × | × | |
| *Exit choice* | × | × | × |
| *Waiting time* | | × | × |
| *Eye-tracking* | × | × | |

### *4.1. Pre-Evacuation Time*

Figure 5 presents how long it took for participants in the HMD experiment and in the physical experiment to leave the room once the alarm was activated. The average pre-evacuation time for participants in the HDM experiment was 32 s. In the case of the physical experiment, the average time was 35 s when outliers were removed. One participant in the physical experiment did not leave the room until 725 s after the alarm, when they were told to do so by the researcher, ending the experiment. If that participant is to be included, the average pre-evacuation time is 52 s. An independent sample two-tailed t-test was performed to compare the pre-evacuation time for both samples. This test resulted in a p-value of 0.2249, which is not statistically significant. Therefore, the analysis suggests that there were no significant differences between the pre-evacuation times obtained by the two methods.

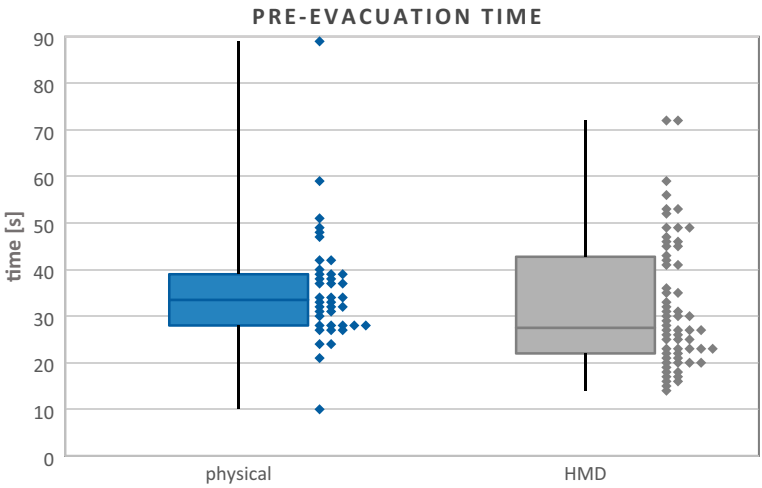


**Figure 5. Time it took for participants in the HMD experiment and the physical experiment to leave the room once the alarm went off.**

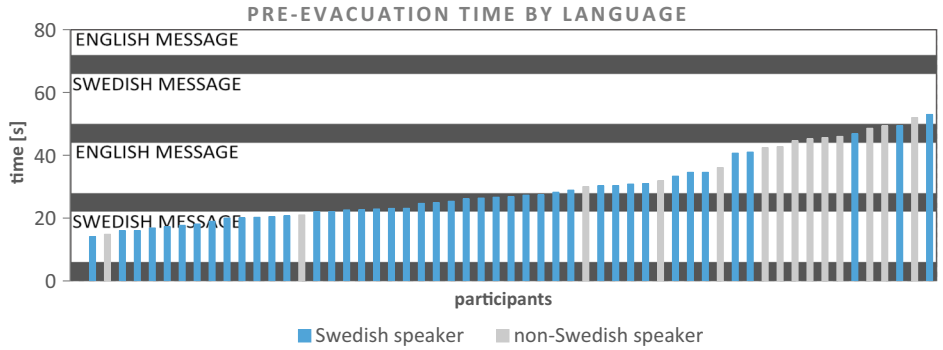


**Figure 6. Pre-evacuation time in the HMD experiment by language. The gray bars represent the tone segment of the alarm. The estimation whether the participant spoke Swedish or not was based on information they gave of their upbringing and may be inaccurate.**

When looking only at the participants in the HMD experiment, a trend can be seen when classifying participants not by scenario but by country of origin. The questionnaire asked "which country did you grow up in?". Estimations based on their country of origin lead to 41 participants counted as Swedish-speakers, and 21 as non-Swedish-speakers. It needs to be emphasized that growing up in a different country does not mean the participant does not speak Swedish. Nevertheless, the assumption was made in order to have a rough estimation on the delay in the evacuation due to not being able to understand the first part of the voice alarm, which was given in Swedish.

With the estimation of whether each participant was able to understand the Swedish message, the pre-evacuation time was plotted again (see Fig. 6). As mentioned in Sect. 2.1, an entire cycle of the alarm lasted around 43 s. In Fig. 6, the gray bars represent the tone segment of the alarm (around 6 s), and the white bars represent the voice message in the corresponding language (around 16 s long voice message in each language). Participants in the physical experiment were not included because all of them were fluent in Swedish. As shown in Fig. 6, the participants leaving the room the earliest were mostly Swedish speakers. The median for the Swedish-speakers was 24.9 s, while the median for the non-Swedish-speakers was 45.6 s.

### *4.2. Noticing Available Escape Routes*

Participants in both the HMD experiment and the physical experiment were asked "Once you left the hotel room, did you notice there were two escape routes?". The question was a multiple choice one, and the available answers were "yes", "no", and "other". The "other" option required them to elaborate. Only five participants answered "other". From those, four gave an explanation in the lines of "not at first" or "not immediately". Those answers were classified as "no", given the fact that the question asked specifically once they left the room, and the participants realizing there were two later on indicates that at the moment they left

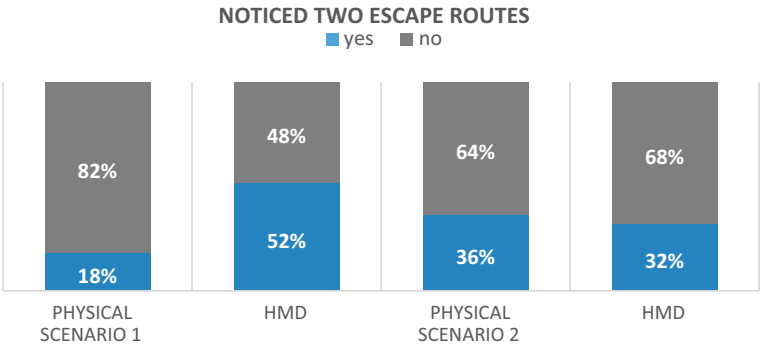


**Figure 7. Proportion of participants who claimed they noticed there were two escape routes once they left the room.**

the room the answer was "no". The fifth one wrote "no", being their answer labeled then as a "no" reply. Figure 7 shows the answers given by the participants for both scenarios and both experiments. The physical experiment did not include the option "other".

A statistically significant difference between the physical experiment and the HMD experiment was observed in the case of Scenario 1, with a $p$-value of 0.0209 and a significance level of 0.05. No significant difference was observed in the case of Scenario 2.

### *4.3. Walking Paths*

Examples of walking paths of the participants in the HMD experiment are presented in Figs. 8 and 9. Most paths were observed in both scenarios, but the ones presented on the figures show four interesting ones. Participants in Scenario 1, more often than not, did not go to the dead end (at the right-hand end of the corridor as shown in Fig. 1) once the alarm went off. It is likely that they either went there before the alarm, or at least looked at it before entering the room. As shown by the dotted line in Fig. 8, some went there. From six participants going to the dead end first, four went there straight out of the room, without turning their heads towards the rest of the corridor. The other two took a glance at the rest of the corridor but decided to go the dead end first. Once in the dead end, all six realized there was nowhere to go from there, they went to the elevator lobby. From there, they could wait for an elevator as described before, or decide to go for the stairs.

Eight participants from Scenario 1 followed the path indicated by the full line in Fig. 8. These participants initially passed by the elevator lobby, and turned back a couple of meters away. From the analysis of the videos, it became apparent that these participants saw the emergency exit signage but did not recognize the door, turning back once they understood they missed it. Once in the elevator lobby, they either waited for an elevator until the end of the experiment or decided to go for the stairs.

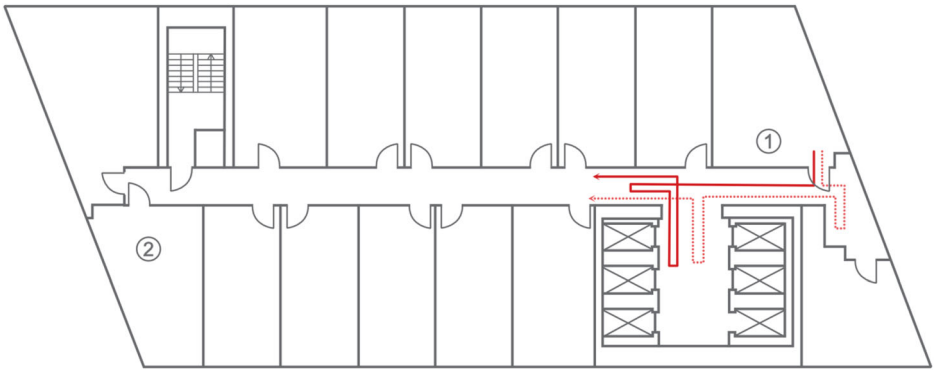


**Figure 8. Examples of common walking paths of participants in Scenario 1. The arrows indicate that the path continues to the stairs.**

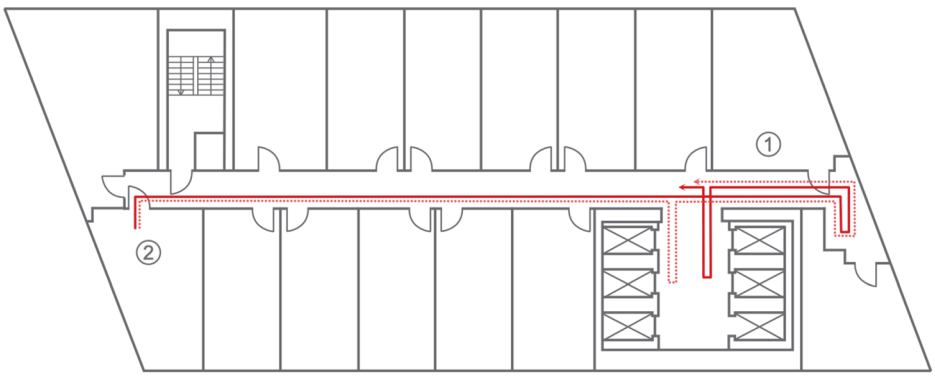


**Figure 9. Examples of common walking paths of participants that were only observed in S2. The arrows indicate that the path continues to the stairs.**

In Scenario 2, some participants followed the paths presented in Fig. 9. Four of them passed by the elevator lobby all the way to the dead end and then went to the elevator lobby, as shown by the full line. Once in the elevator lobby, they waited for the elevators until the end of the experiment or decided to go for the stairs. The dotted line indicates the behavior of four participants in Scenario 2 who went straight for the elevator lobby and waited for the elevators there. After a while, they exited the elevator lobby and went to explore the dead end. Then they either returned to the elevator lobby to wait longer, or went for the stairs.

The seven participants in the HMD experiment, who chose the stairs as first choice, walked in four different paths. In Scenario 1, one participant followed the green line in Fig. 8, and one went directly to the staircase. In Scenario 2, one participant went to the lobby first, and took a quick look at it, but did not approach the elevators or lingered there and then went for the stairs. The remaining four went straight for the staircase.

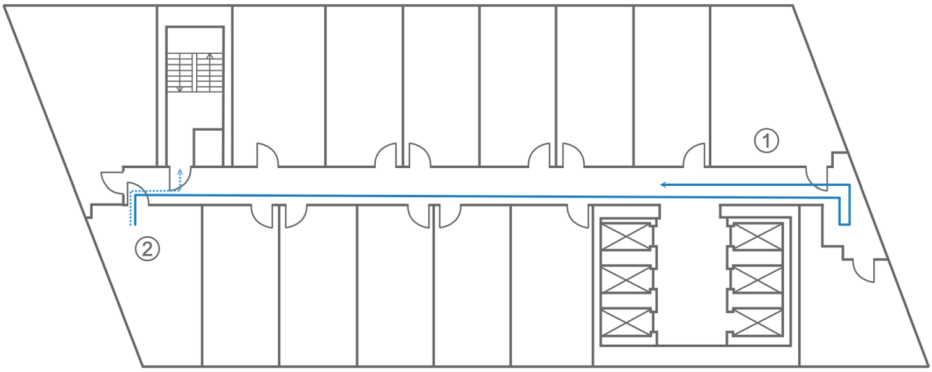


**Figure 10. Walking path participants of the participants using the stairs in the physical experiment.**

The walking paths could be compared to the paths of the two participants who chose the stair in the physical experiment, which are shown in Fig. 10. Both occurrences happened in Scenario 2. One participant went directly to the stairs (dotted line), and that path was also observed in participants in Scenario 2 in Experiment 3. One participant went all the way down the corridor and then turned back and went to the stairs (full line).

### *4.4. Exit choice*

The exit choice by experiment is presented in Fig. 11. The results include only scenarios 1 and 2 in the corresponding experiments, with the exception of the CAVE experiment, which had no scenario 2. In the CAVE experiment and the HMD experiment elevator waiting times were recorded, which were not able to be collected in the physical experiment. Participants in the virtual experiments, who chose the elevators as means of egress, were left to wait up to 5 min in the HMD experiment, and up to 20 min in the CAVE experiment.

The HMD and the physical experiment had similar proportions of participants choosing each means of egress (either ''stairs'' or ''elevator'', which lumped together ''elevator only and ''switched to stairs''). A Fisher's exact test performed showed no statistically significant difference between the two samples at a significance level of 0.05. Therefore, no difference was found between the two samples.

The data on exit choice by scenario is presented in Fig. 12. Notice that only one scenario of the CAVE experiment is included in this paper, and therefore the data on the CAVE experiment is the same in Fig. 11 and Fig. 12. In all cases, there was a preference for the elevators as exit choice. The CAVE experiment had a larger proportion of participants choosing the stairs than the other two experiments. The HMD experiment and the physical experiment showed again similar results to each other. A Fisher's exact test showed no significant difference at a significance level of 0.05. A significant difference was found between the results of the CAVE experiment and the other two.

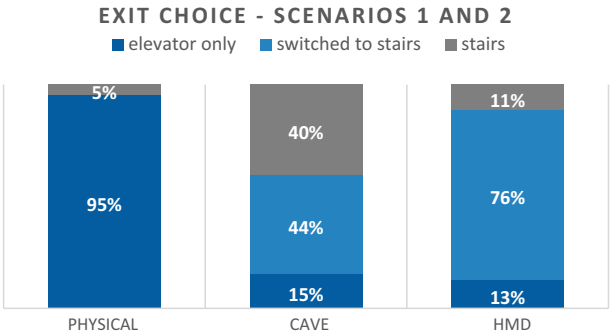


**Figure 11. Proportion of participants using each of the means of egress in the three experiments, independent of scenario.**

### *4.5. Waiting Time*

The waiting time for the elevators in the HMD experiment can be compared to that in the CAVE experiment. Only 18 data points from the CAVE experiment are included, as those are the ones who waited for 5 min or less. Those 18 data points represent around 58% of the sample in Scenario 1 in the CAVE experiment. The results of the HMD experiment showed that only four participants (around 6%) waited for 5 min, which means that the proportion of participants waiting longer than 5 min would have been 6% or less. Figure 13 shows the cumulative proportion of participants waiting times for both experiments.

### *4.6. Eye-Tracking*

The eye-tracking data presented here refers to three objects of interest: the emergency exit sign pointing at the elevator lobby, the emergency exit sign on the elevator lobby door (visible only once the doors closed by the triggering of the alarm), and the emergency sign pointing at the stairs.

Figs. 14 and 15 compare the proportions of participants looking at the emergency signage, classified by experiment and by scenario. A Fisher's exact test performed for the data corresponding to each of the three objects of interest showed no statistically significant difference between the experiments for Scenario 1. However, for Scenario 2 a statistically significant difference was found for two objects of interest: the sign pointing at the elevator lobby, and the sign on the elevator lobby door. Table 5 presents the results of the statistical analysis performed.

### *4.7. VR-Related Results*

Given the intrinsic differences between VR and reality, it is necessary to consider the perception of realism from the side of the participants. A scenario that is not perceived as realistic may not compel the participants to act as they would do in the same scenario in the real world. The following subsections refer only to the HMD experiment, and provide an insight on how participants felt during the experiment, according to the answers they gave in the questionnaire after the

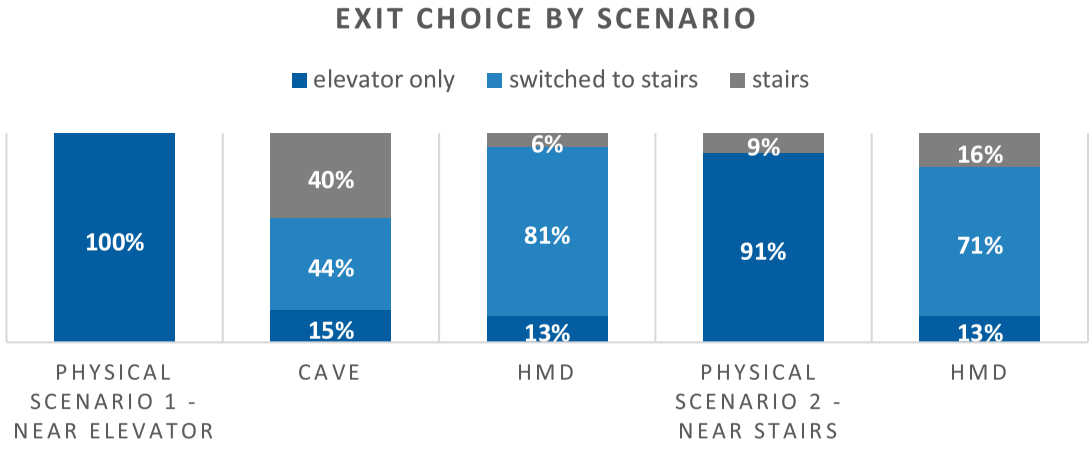


**Figure 12. Proportion of participants using each means of egress by scenario. The CAVE experiment did not run Scenario 2.**

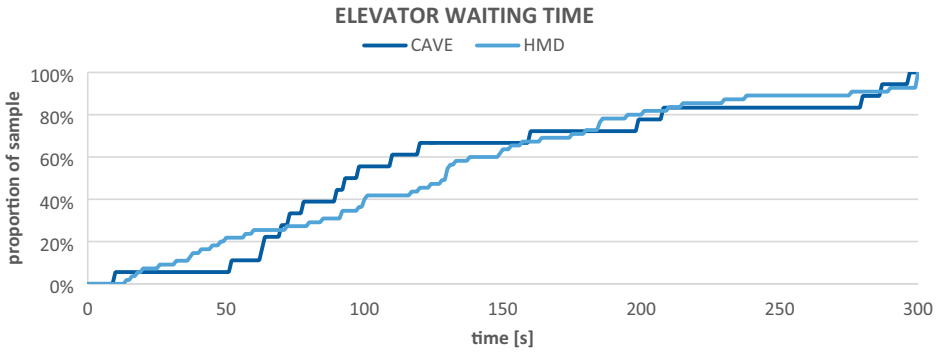


**Figure 13. Total time participants in the HMD experiment and the CAVE experiment waited for the elevators to come. The plot only shows participants who waited up to 300 s (5 min) in the CAVE experiment.**

experiment. Their experiences may help to understand how they perceived the scenario they were exposed to.

*4.7.1. Realism* Participants were asked to rate the realism of different components of the virtual environment by using a Likert scale. As it was indicated in the questionnaire, 1 meant a low level of realism, and 7 a high one. The components they were questioned about were the hotel lobby (which is where they started the experiment and did the calibration of the eye trackers), the elevator lobby on floor 16, the hallway of floor 16, and the hotel room they were assigned to. It should be noticed that both rooms were identical. Figure 16 shows the aggregated ratings given by the participants. The ratings given by the participants show an overall medium–high level of realism, with more than 70% of the participants giving ratings between 4 and 6 in each case.

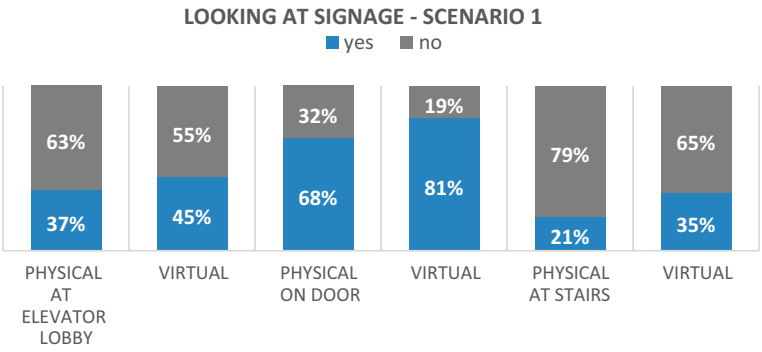


**Figure 14. Proportion of participants per experiment in Scenario 1 looking at the emergency exit signage.**

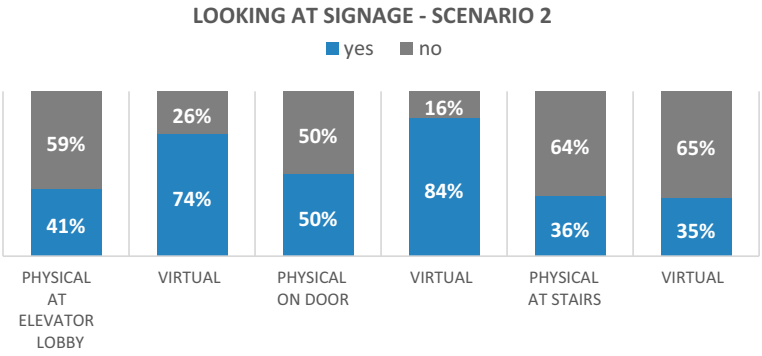


**Figure 15. Proportion of participants per experiment in Scenario 2 looking at the emergency signage.**

**Table 5**
**Results of the Fisher's Exact Test Performed on the Eye-Tracking Data for Each Object of Interest, per Scenario**

| | | | *Look at sign* | | |
|---|---|---|---|---|---|
| *Object of interest* | *Scenario* | *Experiment* | *Yes* | *No* | *p-value (0.05)* |
| Sign pointing at elevator lobby | 1 | Physical | 7 | 12 | 0.7685 |
| | | HMD | 14 | 17 | |
| | 2 | Physical | 9 | 13 | 0.0226 |
| | | HMD | 23 | 8 | |
| Sign on elevator lobby door | 1 | Physical | 13 | 6 | 0.4963 |
| | | HMD | 25 | 6 | |
| | 2 | Physical | 11 | 11 | 0.0142 |
| | | HMD | 26 | 5 | |
| Sign pointing at stairs | 1 | Physical | 4 | 15 | 0.3513 |
| | | HMD | 11 | 20 | |
| | 2 | Physical | 8 | 14 | 1 |
| | | HMD | 11 | 20 | |

*4.7.2. Immersion* The question on immersion asked "Were you immersed in the VR world? (did you "forget" that you were in a laboratory instead of a hotel?)". The answers were presented as a multiple choice with three options: yes, no and other. The "other" option allowed them to elaborate. Figure 17 shows the total number of participants choosing each of the possible answers. There was no difference in the number of participants giving each answer. However, it should be noticed that the "other" option represented an intermediate level of immersion (examples of answers being "half", "partly", "almost", and some participants bringing up that the cable connecting the VR headset to the computer broke the illusion).

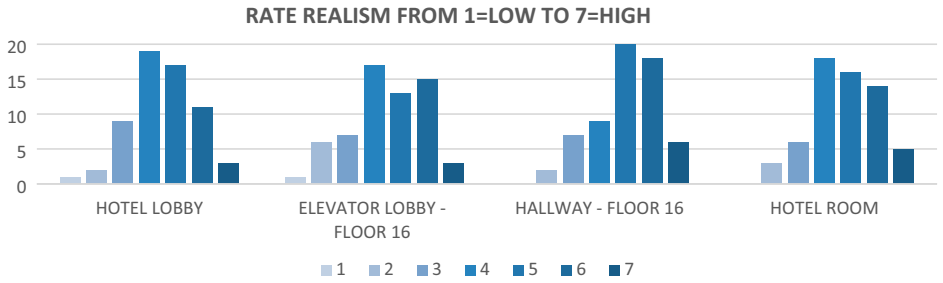


**Figure 16. Number of participants giving each rating of realism to each of the main components of the virtual environment.**

*4.7.3. Sensations* Participants were asked to rate the sensations they experienced during the experiment. A Likert scale was used, indicating 1 as "low" and 7 as "high". Figure 18 presents the number of participants choosing each possible rating for the sensations of insecurity, stress, fear, disorientation, dizziness or nausea and eyestrain. Figure 18 shows the aggregated ratings given by the participants. Overall, the ratings were in their majority between 1 and 4, that is the low to medium range. Around 70% of the participants rated the sensations in this range.

*4.7.4. Equipment* A question on the easiness of using the VR equipment was asked, as an indicator of the comfort of the participants operating in VR. Figure 19 shows the aggregated ratings given by the participants. The majority of participants gave a rating between 2 and 5. It shows an easy to intermediate level of difficulty using the equipment.

*4.7.5. Behavior* One of the last questions in the questionnaire participants filled after the experiment asked "What do you think you would do differently if this scenario happened in the real world?". The question meant to get information about how the participants felt their behavior would be different in a real emergency scenario, and it required a free text answer. Over half of the participants (33 out of 62) answered they would have gone straight for the stairs, without trying the elevators first in a real world situation. Eight participants claimed they would not have done anything different. Others gave less frequent answers, like "would have followed/talked to other people" (four people), "would have waited in the room for the firefighters" (one person), or "would have panicked more" (three people). Although few mentioned panic at all, no panic was observed in the experimental room.

After finishing the experiment, some participants manifested verbally that they only tried the elevators because "the voice" told them to, and that in a real scenario they would not have done that. Other explained they did not think there were any stairs in the virtual environment at all. These two justifications, however, were also brought up by participants in the physical experiment.

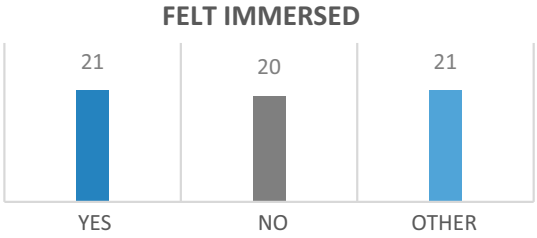


**Figure 17. Number of participants giving each of the three possible answers to the immersion question.**

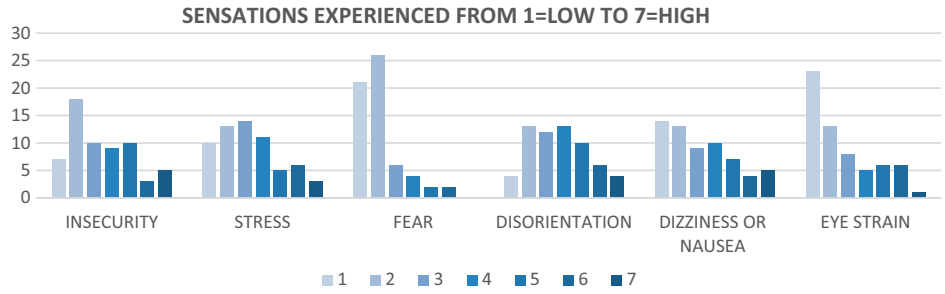


**Figure 18. Number of participants giving a certain rating for the different sensations experienced during the VR experiment.**

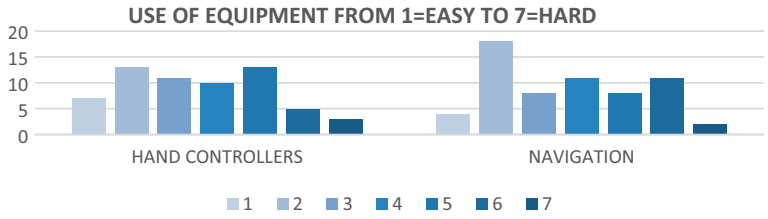


**Figure 19. Number of participants giving a certain rating to the easiness of using the VR equipment during the experiment.**

## 5. Discussion

The results show that the data produced in the HMD experiment was similar to that produced in physical experiment in several ways, and did not contradict the waiting times observed in the CAVE experiment. Nevertheless, in some cases differences were observed between the different datasets. The results will be discussed in the following subsections.

### 5.1. Pre-Evacuation Time

The analysis of the pre-evacuation time of the participants in the physical experiment and the HMD experiment showed a general similarity. However, the similar outcome may be due to different factors that may have been speeding up or slowing down the pre-evacuation time for the participants in each experiment. In the case of the HMD experiment, language may have slowed down the reaction time of the participants who did not speak Swedish, compared to the performance of participants in the physical experiment, which were all fluent Swedish. The non-Swedish speaking participants in the HMD experiment had to wait until the message was played in English, delaying their reaction time for as long as it took the English message to play. The effect of timing and language can be observed in Fig. 6, as the participants leaving the room the earliest were mostly Swedish-speakers.

On the other hand, the different setups from the participants' perspective in each experiment may have made those in the physical experiment more prone to hesitate whether to evacuate or not, compared to those in the HMD experiment. In the physical experiment, they knew they were part of a study focusing on analyzing interior design aspects of the hotel room assigned to them. That "design experiment" was then unfortunately interrupted by an unexpected fire alarm in the building. The participants had to deal with the emergency by themselves, unable to tell if the alarm was part of the experiment or not, being the researcher in charge waiting for them in the ground floor, as far as they knew. In the HMD experiment, the situation was different. Participants were told they were doing an assessment of realism in VR (which was considered somewhat analogous of the analysis of interior design in the physical experiment), and were then interrupted by the fire alarm. These participants knew that such an emergency in a VR experiment was part of the experience, and therefore was clearly part of the experiment.

Moreover, the participants in the HMD experiment were being monitored by a researcher in the experimental room (although not represented in the virtual environment), they knew they were being observed. It is not clear if the presence of the researcher in the room prompted participants to react quickly, but it is possible. One study [6] has shown that actual presence in the experimental room may entice participants to observe social norms that would likely be neglected if no one was around. That study, however, referred only to yawning, which may be a reflex and therefore be based in different mechanisms than those leading to the decision to evacuate. The voice alarm was clear on its instruction to evacuate immediately. Evacuating once a fire alarm goes off is the expected behavior, and ignoring the alarm could be seen as reprehensible. In this situation, the normative social influence could be playing a role, and the participant is prompted to conform to avoid being judged negatively. This hypothesis can be seen as opposite of the results of experiments performed on reaction in a group compared to reaction as a lone individual. However, it should be considered that the participant in the HMD experiment was a lone individual, and had no reason to expect an intervention by the researcher in the virtual environment.

Considerations need to be made when looking at the time it took participants to leave the room. Some participants in the HMD experiment took longer due to dexterity problems when trying to open the door to leave. Others had problems reading the evacuation plan in their room, taking them longer to understand the layout and their initial location. This could be due to effects of the lenses in the head-mounted display (images are sharper in the center of the lenses) or overall low resolution of VR in the case of small objects (like the small ''you are here'' text on the evacuation plan). However, in the physical experiment some participants had to gather their belongings, even put their shoes on. Participants in the HMD experiment had no belongings to collect. The reasons behind the delays in each case may be different, but they may compensate each other.

Other than affecting the time it took for participants to perform certain actions, dexterity using the hand controllers became relevant in the HMD experiment. There were several doors for the participant to operate in the scenarios: the hotel room door (to enter and to leave), the bathroom door, the elevator lobby door and the staircase door. All of them were swinging doors, being the bedroom one and the elevator lobby one also self-closing. From early on it became apparent that many participants struggled with operating the doors, despite the fact that the doors worked as expected in the experimental design. This struggle lead to frustration and impatience in the participants. To solve this problem, collisions between the doors and the virtual body were artificially removed. This means that while the doors remained operational, it was possible for participants to walk through them if they struggled to operate them. This simplification was brought up among the instructions given to the participants before the beginning of the experiment. Some participants could successfully operate the doors. Nevertheless, this improvised solution, although unrealistic, was better at representing the amount of time and effort an adult needs to operate a door in a real world setup.

### *5.2. Noticing Available Escape Routes*

It is unclear what caused the difference between the physical and the HMD samples in the case of Scenario 1. Nevertheless, the perception of available escape routes showed that in almost all cases, the majority of participants claimed they did not notice the two available escape routes. The experiment showed though that independently of the participants' expectations, they are likely to search for an alternative escape route. This could mean that the emergency signage in the hallway was not visible enough for them to notice them and take a moment to integrate that information in their mindset. At the same time, there is no way to check the reliability of the self-reported answers the participants provided in this case, as the information could have come from different sources: seeing all the emergency signage, having read the evacuation plan, assuming there will be two based on evacuation training or other background knowledge, among other.

### *5.3. Walking Paths*

The walking paths walking paths did not show clear trends that could be compared between the physical and the HMD experiment. The diversity in walking

paths showed by participants in both experiments indicates that the behavior of participants in the HMD experiment was not largely different than that of participants in the physical experiment.

In the HMD experiment it was observed that some participants seemed to have missed the elevator lobby doors at first, but corrected their course quickly. Those participants may have had difficulties differentiating between the white door leaves and the long white wall they were on. The sensory affordances of elevator lobby doors may not be ideal considering them as an important part of the evacuation route [18]. Other than them being the same color (or at least very similar), the textures of the materials used in the virtual environment may not have been distinct enough for participants to differentiate them. Applying the right textures may be especially relevant in the case of different materials with similar colors in VR.

### 5.4. Exit Choice

The results show that the exit choice data produced in the HMD experiment was highly similar to that produced in the physical experiment. There was a clear difference with the exit choice in the CAVE experiment. That difference could be due to the equipment used in these experiments. The participant in the CAVE needed to operate the hand-controller in order to turn in VR exclusively, and could not do the instinctual reaction of turning their body. Alternatively, in the HMD experiment, for the participant to turn in any direction, all they needed to do was to rotate their body in that direction and move forward with the hand-controller. It is possible that this better agreement between the physical body and the virtual body helped participants to move in an easier and more natural way. If the equipment affects the way participants move in the virtual environment, future research should focus on exploring its effect and provide insight on the suitability of possible solutions or alternatives to minimize it.

No significant difference was observed in the behavior of participants of the HMD experiment and the physical experiment based on the scenarios that were studied. While a few participants went straight for the stairs in Scenario 2, even fewer (if any) did so in Scenario 1. The statistical analysis showed no statistically significant difference between the samples, so it cannot be concluded that proximity to the means of egress played a role.

### 5.5. Waiting Time

The maximum waiting times in the CAVE experiment and the HMD experiment were different, but the results from the HMD experiment did not necessarily contradict that of the CAVE experiment. Since most of the HMD participants waited less than those five minutes, their absolute waiting time was much shorter than that of participants in the CAVE experiment. Nevertheless, in the CAVE experiment around 58% of the participants left within the first five minutes. The results of the HMD experiment would have been considered contradictory if a large proportion of the participants waited for the established maximum of five minutes. The reason why the waiting time was shorter for the participants in the HMD experiment remains unknown.

As mentioned before, it was not possible to collect waiting times in the physical experiment. This exemplifies how sometimes VR can produce data that may not be possible to collect in a physical experiment. In this case is obviously not possible to validate the VR data by comparing it to a real-world dataset, but the case highlights the advantages of a thorough validation of the VR method.

### *5.6. Eye-Tracking*

The analysis of the eye-tracking data showed that participants in the HMD experiment looked more times at the evacuation signage than participants in the physical experiment did. In Scenario 2, more participants fixated their sight on the emergency exit sign pointing at the elevator lobby and the sign on the elevator lobby doors than in the physical experiment. It is unclear why this difference occurred. However, two possible explanations are presented here. First, as it can be seen in Fig. 3, the emergency signage was easier to spot in the HMD experiment. The increased brightness of the signage may explain why more participants in the HMD experiment looked at this signage compared to the physical experiment. If that is the case, especial care needs to be put when reproducing a real environment in VR, to make sure the participant's perception of the environment does not differ too much. A first approach could be to conduct an environmental assessment, such as a Semantic Environmental Description [13], in order to identify differences in the assessment by groups of people.

The second explanation could be spatial orientation in VR. Given the layout of the floor they were in (largely a straight corridor), participants in the physical experiment would not need to look for the emergency signage in order to find the elevator lobby. Even in the case of a non-emergency scenario, it is unlikely that an adult in one of those hotel rooms would need to follow the signage to find their way back to the elevator. Their basic spatial orientation and the simple layout of the corridor being easily stored in their short-term memory may have been enough to find their way back to the elevator lobby. However, the elevator lobby doors being then closed may have hinder their ability to find the elevator lobby, as they were not closed when they walked in. Nevertheless, spatial orientation is reduced in VR [20, 17], and probably more cues were needed for participants to find their way back to the elevator lobby.

Nevertheless, the differences in the software used to analyze the eye-tracking output may also have played a role in the results. It is likely that there were considerable differences between the eye-tracking software in the two experiments. The commercial software used for analyzing the data output for the eye-tracking glasses in the physical experiment used closed-source code based on proprietary algorithms. The software used in the HMD experiment was a combination of an open-source project for capturing eye-movements together with code developed by the authors to interpret objects of interest, which used a straightforward algorithm based on the defined length of fixations. It is possible that the proprietary code considered some factors that the code developed by the authors did not, resulting in a discrepancy on the way the eye-tracking data was measured. The discrepancy

could be minimized by using the same source code in the eye-trackers used in both experiments.

### 5.7. VR-Related Results

The VR-related results showed a positive outcome in terms of realism. Most participants gave ratings between four and six out of seven. It should be noticed though that the highest score (seven) was not used very often. However, on average the assessment of realism was overall acceptable, as only ca. 30% of the participants gave low scores. Reliability of these results also come into question, especially since participants were not given a parameter for what realism means. For some it could mean ''exactly like reality'' while for others it could be closer to ''good enough for a computer game''.

Participants were asked about their level of immersion, although the intention was to ask about presence. Nevertheless, the guidance they were given (''did you forget you were in a laboratory instead of a hotel?'') was probably better understood by them than the technical term on itself. The immersion/presence participants claimed they experienced was equally distributed between the three possible answers (i.e. ''yes'', ''no'', and ''other''). This outcome can be considered neutral, as the same number of people argued on each end of the spectrum, and the ones in between were not decisive. However, the impressions of those who claimed had an intermediate level of immersion/presence indicate that their immersion level could be improved. The cable that connects the HMD to the computer may be replaced by a wireless solution. Alternatively, VR equipment that integrates all of the processing hardware into the HMD itself does exist. That type of equipment, however, may present limitations in the size or complexity of the VR scenario to be run, as its processing power is lower than that of a high-end gaming computer.

The rating participants gave to the different sensations they experienced during the experiment, as shown in Fig. 18, showed participants were not too distressed by the scenario they were exposed to. Fear got the lowest ratings, and no participant gave the maximum rating for this sensation. Stress was relatively low, with ca. 2/3 of the ratings given between one and four. It could be argued that higher stress levels would be expected in such an evacuation. However, no data was collected on stress level in the physical experiment. Moreover, given the fact that the emergency in this scenario consisted of a fire out of the participants' sight, it is not necessary that they would experience high levels of stress. On the other hand, the rating presented here is self-reported, and participants were given no clear instructions on what each of the different ratings entailed.

The claim made by around half of the participants in the HMD experiment, that they would not have used the elevator in a real-world scenario like that, can be challenged by the data gathered in the physical experiment. A match in the exit choice data between the physical experiment and the HMD experiment suggests that even in a real-world scenario the tendency was to use the elevators, at least as first choice. Nevertheless, the HMD experiment was indeed designed to replicate the physical experiment as close as possible, and the disparity between the claims of the HMD experiment participants and the actions of the physical experiment

participants may be another piece of evidence of the lack of reliability of self-reported data. Even though participants in the HMD experiment thought they would not behave in the same way in reality, the comparison with the behavior of participants in the physical experiment shows otherwise. This difference shows that participants had the perception that the HMD experiment was not a fair representation of real-world conditions, when the data shows it was at least very close.

The fact that many participants from both samples also pointed out that they tried the elevators first because "the voice" told them to points out the effectiveness of the voice alarm in this case. The objective of such a voice alarm is to give clear instructions to the building occupants in order to achieve a quick and safe evacuation. The participants' claim that they used the elevators because of those instructions means that the voice alarm was effective. It should be noticed though that the voice alarm did not literally instruct them to use the elevators. It simply stated that it was possible to do so. Participants' interpretation of the message as an instruction was of their own making.

### *5.8. Limitations and Future Work*

This study meant to assess similarities and differences between the three experiments given the difference in the research method and equipment used in each. By no means can the data presented here be applied in other contexts, as it is not representative of real hotel guests.

The three experiments had relatively small sample sizes, which means that small differences among participants could have a large impact in the overall proportions. Larger samples would reduce this effect and offer more conclusive results.

The samples were collected independently for the three experiments, which took place in different years and different cities in Sweden. Ideally the three experiments would have been conducted in parallel, assigning participants in the same pool randomly to each of them. The samples compared in this study were similar in average age (averages between 25 and 33 years old) and gender distribution. The differences between the samples are likely to have introduced bias in the results. However, it is unlikely that those differences affected their behavior in the many aspects compared in this paper (exit choice, pre-evacuation time, walking pattern, looking at signage, etc.) in a meaningful way. Participants also differed in country of origin, being the HMD sample the most international one, while the physical experiment had mostly Swedes. There is no data on nationality in the CAVE experiment. Cultural background could play a role in the decisions made in an emergency. Nevertheless, other than the expected delay in the pre-evacuation time for non-Swedish speakers, the HMD and the physical samples were not too different in terms of exit choice and pre-evacuation time. The walking paths were diverse in both samples, which shows no large difference between the HMD and the physical experiment.

Null-hypothesis significance testing (NHST) may not be the best approach to compare the data. This is illustrated by the selection of a significance level. In this study, the significance level chosen was 0.05. Due to multiple comparisons, a cor-

rection such as Holm-Bonferroni [7] should be introduced. The correction would reduce the significance level to a smaller number. However, for this study, a smaller significance level is an advantage: the smaller the significance level, the lower the chances for a statistically significant difference in the results. No statistically significant difference in the context of this study is a good result. NHST was developed to identify differences, not similarities. This study aims for similarities. No statistically significant difference is no evidence of similarity between the VR and the physical results, but that a level of similarity cannot be ruled out. Therefore, if the significance level was reduced due to multiple comparisons, the study will be more successful at showing no statistically significant differences, helping its own case. NHST was chosen because of limitations in sample size, that would violate assumptions of other, possibly more suitable tests.

This study aimed at comparing the samples obtained in the different experiments. However, a different approach would have been to compare the effect of the scenarios in the different experiments. As an example, the results on exit choice showed that more participants used the stairs in Scenario 2 than in Scenario 1, in both the physical and the HMD experiment. In fact, in both experiments, the difference between the samples choosing the stairs was roughly 10%. It is then apparent that the effect of proximity to the stairs was the same in the physical experiment and the HMD experiment. That effect was not statistically significant, but it was present in both, which is another indication of similarity between the results of the HMD experiment and the physical experiment. Future studies could focus in assessing whether the same effect observed in a physical case is observed in a VR reproduction.

More studies are needed to make an in-depth assessment of the application of VR experiments for Human Behavior in Fire research. The results of the present study point at different directions for future research. The reduction on spatial orientation in VR needs to be studied in detail, in order to incorporate measures to minimize their impact. This topic is especially relevant in way-finding studies in VR, which need to take measures to minimize the difference to produce realistic data.

The visual comparison between the virtual environment used in the HMD experiment and the building used in the physical experiment shows differences in terms of lighting and perception of dimensions. It is clear that the distortions of the camera used to take the picture from the real building play a role, but other aspects such as photorealism, effect of lighting, and perceived distances in VR need to be analyzed. It remains unclear whether a photorealistic virtual environment is necessary to conduct experiments in VR, or a simple, even cartoonish appearance produces the same type of data with lower costs in terms of computational power and design.

Lastly, non-player characters could be included in other scenarios, to study the effect of social influence in both exit choice and waiting times. In these experiments the participants were on their own, while in a real-world evacuation in such a scenario, most likely there will be other people on the same floor. The effect of social influence in this condition needs to be considered, especially if the VR experiment aims at collecting data on waiting times.

## 6. Conclusion

The HMD experiment showed a high level of similarity with the physical experiment in terms of exit choice and pre-evacuation time. The similarities suggest that the collection of that data on exit choice and pre-evacuation time is indifferent to the method used (HMD or physical experiment) in this case. The differences between the samples could have introduced a bias in the comparison of the methods, and the relatively small sample size and to the many factors that make the behavior of individuals unique. Despite the differences between the samples, the results support the use of VR as a research method in Human Behavior in Fire, but they are not conclusive. More research is needed to continue validating VR as a research method, as the present study does not cover all possible VR scenarios and variables to be assessed.

The walking paths were contrasted, but no clear difference could be observed between the HMD and the physical experiment samples. Both samples showed a wide spread of walking paths, making the HMD results consistent with reality.

The differences with the results from the CAVE experiment suggest the equipment used may play a role. It is possible that the locomotion in the CAVE experiment was part of the reasons why the behavior differed significantly, and it could be worth to explore the impact of the equipment used.

The eye-tracking data differed in some cases in the physical experiment and the HMD experiment. The discrepancy may indicate differences in the perception of the virtual environment by the participant, and even differences in the spatial orientation in VR and in reality. More research is needed to determine how the differences could be minimized.

The differences between the choices participants did in the HMD experiment and their answer about what they would do if the experiment took place in reality were interesting but they became remarkable once the behavior was compared to that of participants in the physical experiment. The HMD and the physical samples were similar in their exit choice, despite the participants' prediction. This result highlights that the HMD experiment was able to reproduce realistic behavior, even though the perception of participants was different.

Both VR experiments allowed the collection of waiting times for the elevators, an important aspect to be considered in the design of an evacuation concept based on the use of elevators. Collecting this data in the physical experiment was not impossible, but it was unviable due to the cost of stopping normal business operations for the hotel. This difference in costs for a research project is an advantage for the VR experiment method, and therefore more studies are needed to identify limitations of the method and compensate for differences with reality.

## Acknowledgements

The authors of this paper would like to acknowledge the Swedish Fire Research Board (Brandforsk—grant number 217–171) for funding this study. In addition,

the authors would like to thank Lic. Eng. Kristin Andrée for making data from the CAVE experiment available for this study.

The authors would also like to thank Dr. Håkan Frantzich and Dr. Enrico Ronchi for their valuable contributions to this paper. The authors would also like to thank Dr. Johan Lindström from the Centre for Mathematical Sciences at Lund University for their advice on statistical analysis.

## Funding

Open access funding provided by Lund University.

## Open Access



## References

1. 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 Mater 40(4):554–567. https://doi.org/10.1002/fam.2310
2. 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 Saf J 109:102861. https://doi.org/10.1016/j.firesaf.2019.102861
3. Arias S, Nilsson D, Wahlqvist J (2020) A virtual reality study of behavioral sequences in residential fires. Fire Saf J. https://doi.org/10.1016/j.firesaf.2020.103067
4. Arias S, Wahlqvist J, Nilsson D, Ronchi E, Frantzich H (2020) Pursuing behavioral realism in virtual reality for fire evacuation research. Fire Mater. https://doi.org/10.1002/fam.2922
5. Bode NWF, Codling EAJFT (2018) Exploring determinants of pre-movement delays in a virtual crowd evacuation experiment. Fire Technol. https://doi.org/10.1007/s10694-018-0744-9
6. Gallup AC, Vasilyev D, Anderson N, Kingstone A (2019) Contagious yawning in virtual reality is affected by actual, but not simulated, social presence. Sci Rep 9(1):294. https://doi.org/10.1038/s41598-018-36570-2

7. Holm S (1979) A simple sequentially rejective multiple test procedure. Scand J Stat 6(2):65–70
8. Kinateder M, Gromer D, Gast P, Buld S, Müller M, Jost M, Pauli P (2015) The effect of dangerous goods transporters on hazard perception and evacuation behavior - a virtual reality experiment on tunnel emergencies. Fire Saf J 78:24–30. https://doi.org/10.1016/j.firesaf.2015.07.002
9. Kinateder, M., Ronchi, E., Nilsson, D., Kobes, M., Müller, M., Pauli, P., & Mühlberger, A. (2014, 2014). *Virtual reality for fire evacuation research*. In: Paper presented at the Federated Conference on Computer Science and Information Systems
10 Kinateder M, Warren WH (2016) Social Influence on evacuation behavior in real and virtual environments. Front Robot AI. https://doi.org/10.3389/frobt.2016.00043
11. Kinateder M, Warren WH, Schloss KB (2019) What color are emergency exit signs? egress behavior differs from verbal report. Appl Ergon 75:155–160. https://doi.org/10.1016/j.apergo.2018.08.010
12. Kobes M, Helsloot I, de Vries B, Post JG, Oberijé N, Groenewegen K (2010) Way finding during fire evacuation; an analysis of unannounced fire drills in a hotel at night. Build Environ 45(3):537–548. https://doi.org/10.1016/j.buildenv.2009.07.004
13. Küller, R. (1972). *A semantic model for describing perceived environment*. Retrieved from National Swedish Institute for Building Research, Stockholm:
14. 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 Technol. https://doi.org/10.1007/s10694-020-01046-1
15. Mossberg A, Nilsson D, Wahlqvist J (2020) Evacuation elevators in an underground metro station: a virtual reality evacuation experiment. Fire Saf J 120:103091. https://doi.org/10.1016/j.firesaf.2020.103091
16. Moussaïd M, Kapadia M, Thrash T, Sumner RW, Gross M, Helbing D, Hölscher C (2016) Crowd behaviour during high-stress evacuations in an immersive virtual environment. J R Soc Interface 13(122):20160414. https://doi.org/10.1098/rsif.2016.0414
17. Nguyen-Vo, T., Riecke, B. E., & Stuerzlinger, W. (2017). *Moving in a box: Improving spatial orientation in virtual reality using simulated reference frames*. In: Paper presented at the 2017 IEEE Symposium on 3D User Interfaces (3DUI)
18. Nilsson D (2009) Exit choice in fire emergencies: influencing choice of exit with flashing lights. Dept. of Fire Safety Engineering and Systems Safety, Lund University, Lund, Sweden
19. Ronchi E, Nilsson D, Kojić S, Eriksson J, Lovreglio R, Modig H, Walter AL (2015) A virtual reality experiment on flashing lights at emergency exit portals for road tunnel evacuation. Fire Technol. https://doi.org/10.1007/s10694-015-0462-5
20. Witmer BG, Bailey JH, Knerr BW, Parsons KC (1996) Virtual spaces and real world places: transfer of route knowledge. Int J Hum Comput Stud 45(4):413–428. https://doi.org/10.1006/ijhc.1996.0060