UNIVERSITY OF CAPE TOWN

DEPARTMENT OF COMPUTER SCIENCE

CS/IT Honours Project
Final Paper 2022

Title: The Effect of User Control on The Elicitation of Fear in Virtual Reality

Author: Iain Hutcheson

Project Abbreviation: VRSad

Supervisor(s): Prof James Gain, Dr Gosia Lipinska

Category	*Min*	*Max*	Chosen
Requirement Analysis and Design	*0*	*20*	0
Theoretical Analysis	*0*	*25*	0
Experiment Design and Execution	*0*	*20*	20
System Development and Implementation	*0*	*20*	10
Results, Findings and Conclusions	*10*	*20*	20
Aim Formulation and Background Work	*10*	*15*	10
Quality of Paper Writing and Presentation	*10*		10
Quality of Deliverables	*10*		10
Overall General Project Evaluation (*this section allowed only with motivation letter from supervisor*)	*0*	*10*	0
Total marks		**80**	

The Effect of User Control on The Elicitation of Fear in Virtual Reality

Iain Hutcheson
Department of Computer Science
University of Cape Town
Cape Town, South Africa

HTCIAI001@myuct.ac.za

ABSTRACT

Virtual reality (VR) offers a unique method of fear induction when compared to other affective media, as users can interact with a virtual environment (VE) as if they were really there. This paper investigates the role that this interaction, specifically its restriction, and the resulting loss of agency, plays in enhancing fear elicited in VR. The study used both within-subjects design, to test for overall effects of the VE exposure among the entire participant sample, as well as between-subjects design to test the difference in the level of fear elicited in a treatment where user's control is suddenly restricted (Tool Loss). Subjective measures for fear elicitation included verbal and non-verbal questionnaires for assessing emotion. Objective measures included changes in HR, RR and SCL. In the within-subjects study, these were compared between baseline and post-immersion measures. In the between-subjects study change scores were calculated by taking the average measures of each over the last 40 seconds of immersion, and subtracting the average measures of the preceding period. Participants displayed significant changes in HR and SCL after VE immersion in the within-subjects study, however the subjective results did not indicate that self-reported fear was the primary emotion experienced during immersion. In the between-subjects study, participants in the Tool Loss treatment displayed a significantly greater increase in HR during the monitored period. Verbal subjective measures (DEQ and VAS) did not differ significantly in self-reported fear. Non-verbal measures (SAM) indicated that participants in the Tool Loss treatment experienced a significantly greater feeling of general excitement and a significantly lower feeling of control over their emotions, than the No Loss group. These results suggest that (i) the environment was successful in producing arousal, and possibly fear, (ii) the variable of agency, resulting from a loss of user control, may have had the effect of producing greater sense of fear in participants.

1 INTRODUCTION

The ability of an interactive virtual environment to fully embody a user is unique among affective digital media. Virtual Environments (VEs) use a combination of immersive audio-visuals and Virtual Reality (VR) peripheral hardware in order to "place" a user in a computer-generated world.

The immersive experience that VEs offer has great utility in research involving human behaviour . Since emotions could be said to dictate human behaviour, then their controlled elicitation in virtual reality could be instrumental in understanding how they affect people so significantly. In the context of psychological and social research, VEs allow for the realistic simulation of dangerous, or terror-inducing, scenarios in a way that maintains experimental control and never physically endangers the participant. This is invaluable in the case of Virtual Reality Exposure Therapy (VRET), which has shown positive results for the elicitation and treatment of social anxiety with professional guidance. In VRET treatment for specific phobias, patients must be repeatedly faced with fear arousing stimuli in order to overwrite the memory construct in their brain responsible for producing the fearful behaviour [22]. As such, this relatively safe and realistic method for maintaining experimental control is highly valuable.

At present, there is little research involving actual methods of eliciting emotions in VR, in favour of the more intangible phenomenon of presence. This study aims to explore the effects of user control on fear elicitation. The paper will introduce the design of the system, as well as the two broader experiments.

Research Questions:

(1) Can a VE designed to trigger acrophobia generate a fear-response in a non-phobic sample?
(2) In a fear-inducing VE, does restricting the user's ability to complete some stated interaction task lead to a heightened fear response?

Hypotheses:

(1) Post-immersion psychophysiological measures will indicate higher levels of arousal than baseline.
(2) A triggered event which prevents the user from completing the environment's stated interaction task (Tool Loss) will cause a significant increase in these same psychophysiological markers, when compared to a control group (No Loss).
(3) Subjective measures of emotion will indicate that, during immersion, users experienced fear significantly more than

other emotions. This will be enhanced in the Tool Loss treatment.

2 RELATED WORK

In an interactive VE, the user is given selective control over the simulation in order to allow them to complete some interaction task. These tasks usually involve some combination of object manipulation and navigation, using some tracked form of locomotion, in order to achieve some end goal [14,17]. In a VE designed to elicit emotion, the interaction task may take the form of an action which, itself, produces an emotion-arousing stimulus. For example, in the case of a VR scenario simulating acrophobia, the act of a phobic user looking down from a great height, alone, may result in a significant fear response, as noted by Diemer et al. [2016]. Alternatively, the VE may prompt some scripted action from the environment, intended to stimulate an emotional response. In a VR horror game, for example, where the user must use a gun to fight AI-scripted zombies. The task of manipulating the gun to shoot a zombie may not cause a significant emotional response in isolation, but if the zombie's AI is then triggered to wail or writhe around, this stimulus could cause the arousal of fear, disgust or even excitement in the user [20]. To this effect, control, characterized by interaction tasks that cause emotional responses, can be seen as the activating factor in the elicitation of emotion in a VE. Despite this, empirical research exploring the effects of control, in interactive affective media is limited.

A study by Madsen [2016] suggests that user control enhances fear evoked in a non-immersive horror video game. The experiment tested two groups, one playing through the game with limited control, the other watching the game. The interaction group's fear increased significantly more than the watching group, especially when user control was suddenly restricted. Despite this the only self-reported difference between the groups was perceived level of entertainment, also consistent with the physiological measures found [18] This shows that fear may be heightened when control of a simulation is introduced, and suddenly restricted. This interaction also arose in a study by Lin [2018] in a VR horror game. Users with the highest subjective experience of fear displayed avoidance behaviours indicative of the activation of the fear structure in emotional processing [21]. Importantly, these behaviours were most prominent when users lost control of their weapon and could no longer protect themselves. Here, a loss of user control made users feel unable to cope with the already-frightening scenario, producing overt avoidance behaviour.

This sequence of events where 1) users are given some interaction task 2) their ability to complete the task becomes restricted 3) resulting in a heightened fear response, could be used as a model for increasing fear elicitation in VR. Unfortunately, the few studies that have identified this interaction have both used horror games as their methodology [17,21] which are designed for entertainment. This means that they intend to elicit other emotions in addition to fear, such as excitement and relief [20], so may not prove as experimentally valid as an environment developed for the purposes of an experiment.

3 ENVIRONMENT DESIGN AND IMPLEMENTATION

Figure 1: Aerial views of the cityscape and street scene around the central platform

3.1 Environment Design

3.1.1 Theme and Setting. Whereas previous experiments involving fear elicitation in VR at UCT have utilized thematic horror elements, the primary design focus of this VE was to create a fear-inducing scenario that would still be grounded in realism. Acrophobia, or the fear of heights, was chosen as the central fear trigger, as it has clear psychophysiological signs of emotional arousal, even in non-phobics [3] and has been validated for its effectiveness when simulated in VR throughout past research [15, 3].

An elevated, urban construction environment was chosen as the location for the VE. This would provide several thematic design advantages over other height-based VR scenarios. Firstly, a city scene could be populated with surrounding buildings and street level props, which were thought to be more effective in creating a realistic sense of scale than a sparser mountainside scene [13], for example. This sense of scale could also be aided by having the user interact with a construction lift, which would rise to a platform from the street. Finally, the construction environment would provide the obvious, thematically sound interaction task for the VE, in the form of various machine parts that would need to be fixed with a tool.

3.1.2 Script. The participant starts the VE on a lift platform, which is at ground level. They then press a button on a control panel and are then raised to another elevated platform, approximately 23m from the ground. Shortly after they reach the top, various machine parts on the platform will begin to spark/smoke and make a loud noise. The user is then prompted to exit the central lift and step out onto the surrounding platform. Once on the platform, the user will be told that they must repair the 3 malfunctioning machine parts in order to fix the lift, which drops to a point just below the platform and begins to move erratically, with accompanying particle and sound effects. Once they have achieved some combination of either task completion, or automatically timing out of the portion of the scenario, the lift is prompted to drop to the bottom of the shaft. Following this, the user's vision slowly fades to black, after which the VE has been completed.

3.1.3 Treatments. To assess the second research question, two distinct treatments for the VE were developed. In the control (No Loss) condition, the user's ability to interact with the main task of the VE is not restricted, and they have the ability to fix all 3 parts until they are timed out at 5 minutes. In the Tool Loss condition, users will lose their tool in a scripted event approximately 40 seconds before the end of the simulation. At this point, their ability to complete the interaction task is restricted, and they will be unable to fix the remaining machine part. The behaviour of the lift is identical across both scenarios, so any changes in emotion measured can be determined to have arisen from the agency-restricting event.

3.2 Methodology

User Centered Design – the environment was being built from scratch, important that it be usable, while retaining experimental design validity. Unfortunately, both supervisors were unavailable for in-person requirement setting and testing during the development period. Instead, several participant groups were utilized at various points during the testing period in order to refine various features. This had the added benefit of acting as candid, pilot tests for each design iteration, with participants experiencing the VE with no prior knowledge, as they would in the final experiment itself.

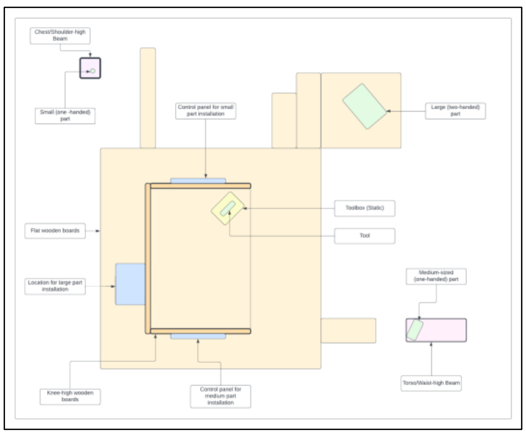

Figure 2: A prototype layout for the environment

3.2.1 Phase 0: Pre-testing. Before starting development in Unity, it was important to draft a layout for the platform where the VE would take place. This would need to fit entirely in the 2.93m x 2.55m dimensions of the UCT Ishango Labs Experiment Room 2's simulation area. Teleportation, a common form of VR locomotion which allows for the navigation of a larger environment through the stitching together of smaller room-sized areas, would not be possible to employ, due to the need for the environment to conform to the configuration of the static haptic ledge.

Feedback from the previous VR honours project indicated that users were likely to drop grabbable objects, preventing themselves from completing the scenario. This would jeopardize the experimental validity of the controlled "tool loss" event. To solve this, a conditional respawn on the tool (preventing it from being lost prematurely) was developed and all machine parts for the user to repair were fixed to the platform.

3.2.2 Phase 1: *Baseline Functionality*. In the first phase, a mock environment was developed with simple functionality for the lift rising via button press, as well as the various fixable part interactions. This was done using low-polygon placeholder models. Testing in this phase simply involved asking participants whether they experienced any motion sickness while ascending on the lift, and if the platform had enough space at the top to move around the platform without fear of falling off (in the VE) or hitting the walls (in real life).

3.2.3 Phase 2: Improved Fear Events. The day-night lighting cycle was a major part of the initial development process. This included custom scripts to adjust the ambient and directional lighting intensity and colour over time, as well as a custom shader to blend between a day and night skybox dynamically. In a requirements-setting meeting, our supervisors indicated their concern that this alone would not produce an urgent enough scenario, and that the distraction of the cognitive task may detract from the subtle change in visuals over time, potentially eliciting frustration, rather than fear, when agency is restricted. The addition of higher polygonal building models also had the negative effect of adding very large rendering times to the mixed lighting system, which used a combination of baked and dynamic shadows. These sometimes did not have time to render back to the starting state of the environment between simulations and would appear pitch black despite bright ambient lighting values. To create a sense of danger, further malfunction behaviour was added to the lift, which would worsen as the VE's state progressed over time.

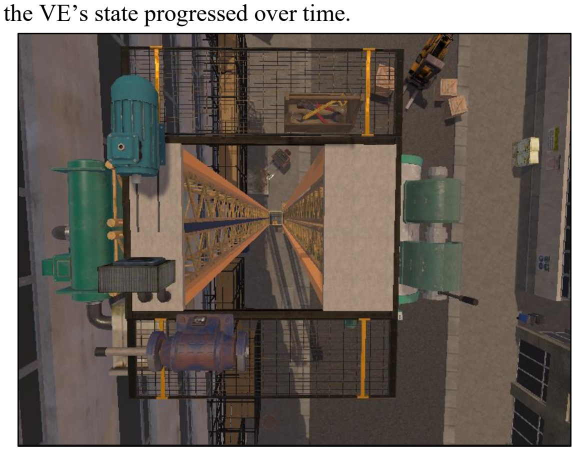

Figure 3: The final platform layout

3.2.4 Phase 3: Tutorial Implementation and Usability. The final development phase involved the addition of a tutorial, covering the interaction task of the VE, as well as various UI elements to make the task clearer to users. High load times between the tutorial and the final environment, caused by the lighting rendering, meant that users in the pilot tests would often forget what the required tasks if the VE were before getting to the lift platform. This would have had a disastrous effect on the capacity of the VE to successfully elicit fear. Some users in the pilot tests were so confused by the task that, after the initial shock of ascending the lift was over, they simply idled around the lift platform, reporting diminished fear. Finally, in light of this, it was decided to scrap the dynamic lighting altogether.

3.2.5 Tools. The VE was developed in the Unity game engine, which utilizes C# as a scripting language. VR functionality was integrated with the use of Unity's XR Framework plugin as the Software Developers Kit, as well as the XR Interaction Toolkit for basic object interactions. Unity's Universal Rendering Pipeline (URP) plugin was used for post-processing effects, including anti-aliasing. The High-Definition Rendering Pipeline (HDRP) is better for visual fidelity; however, its rendering speed is lower, and it is more GPU intensive at runtime, which can throttle frame rate. The decision to use the URP was motivated by the literature, which indicates that graphical fidelity is non-essential in creating a VE that is effective in eliciting emotions [1]. Low framerates, however, have been shown to have significant negative effects on the capability of a VE to elicit emotions [2].

3.2.6 Hardware. The VE was developed using a HTC Vive Pro VR Rig, with associated HMD and controllers. This was carried out on a PC with a GeForce GTX 3060Ti graphics card.

3.2.7 Haptics. The environment utilised a configuration of wooden boards (of height 2.4cm) to act as a haptic ledge below the user. This added a tactile response when users placed their foot at the edge of the elevated platform, shown to enhance immersion [Madsen]. This was helpful in preserving immersion when the lift malfunctioned and dropped, as the central platform was slightly lower than the surrounding boards (1.2cm high), meaning that the user had to skirt around the sides rather than walking straight over.

Figure 4: The haptic ledge

3.3 Interaction Task

When designing the interaction task(s) of the VE, it was important to consider that the user's perception of their ability to complete the task would act as the main difference between the two interaction treatments. This means that the tasks would need:

(1) To be simple enough for first-time VR users to understand
(2) To be complex enough to maintain the user's belief that they are in a dangerous, construction environment.
(3) A clear start and end point for user interaction to indicate progression through the scenario

The second of these criteria is important for the fostering of a psychological phenomenon known as Plausibility Illusion (Psi) in the VE, which ties user immersion to the belief that one's actions (in the VE) are occurring in reality [12]. Psi, as well as the broader phenomenon that it comprises part of, called presence, has been shown to have a significant effect on the elicitation of fear in VR, throughout past research.

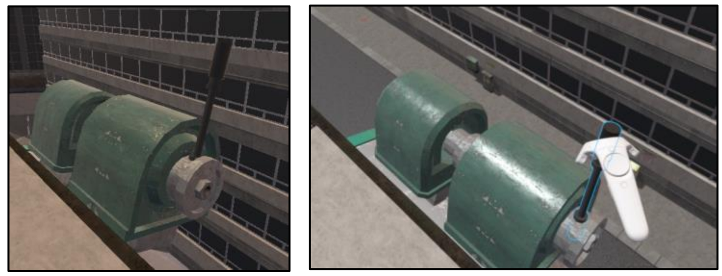

Figure 5: Example of the highlighted movable grab interaction object

3.3.1 Fixable Machine Parts. The VE featured three distinct machine parts, which the user was required to repair in order to complete the scenario. Each fixing task had a clear starting and ending point. Visually, this took the form of the emission of various particle effects, such as smoke or sparks, to indicate that the part was broken. This was also reflected in the VE's surround audio, where each part had distinct sounds for their broken, half-fixed (via movable mechanism) and fully fixed (via nut and bolt) states. Another visual indicator for interaction was the highlighted grab interactable, which the user was required to move into place in order to fix the movable mechanism.

To fix each part they must, in the following order:

(1) Interact with some obviously highlighted moveable part on the malfunctioning machine
(2) Use a tool, spawned from the toolbox, to fasten a bolt in the nut and bolt mechanism, placed somewhere on the part

3.3.2 The Toolbox. This object allows the user to spawn a tool directly into their hand upon simple trigger press with the controller. It interacts with an invisible tool respawn plane far below the platform to detect whether the spawned tool object has

Figure 6: The toolbox object with interactable tool in place

been dropped off the side. The box will only spawn another tool after the respawn plane has been triggered. In the tool loss treatment, after the loss event has been triggered, the box will no longer respawn tool objects.

3.3.3 The Tool Interactable. The tool object has the appearance of a yellow wrench. This is used to interact with the nut and bolt mechanisms in the second, and final, stage of each interaction task.

The tool has realistic physics interactions with the rest of the environment and plays a light collision sound when dropped to the platform. In the Tool Loss scenario, this is forcibly removed from the user's hand in order to create a loss of agency.

3.3.4 Nut and Bolt Mechanisms. The nut and bolt mechanism is the final fixable component of each machine part. Each one can only be triggered after the respective moveable mechanism for the part has been fixed, after which they become active for interaction. When a tool object is held nearby an active nut and bolt mechanism, a bolt will appear between the tool and the nut. Upon activation, using a simple button press, the bolt is fastened with an accompanying animation and sound effect. The nut and bolt mechanism then communicates with a central event handler to advance the state of the simulation, worsening lift malfunction behaviour.

Figure 7: Example of an empty nut and bolt mechanism as well as a nut with a bolt appearing after task completion and tool interaction

The nut and bolt mechanisms were designed to highlight the importance of the tool, as a key item in the completion of the interaction task, through repetition. When the tool is lost, although the user can continue to try to fix the movable parts, they cannot interact with the nut and bolt mechanisms any further. This creates a loss of agency, as no further repairs can be made, and the task is no longer possible to complete.

3.4 Fear-Inducing Events

Several discrete trigger events were designed in order to elicit fear in the participants while they attempted to complete the stated interaction task of the VE.

3.4.1 Lift Raise. The initial fear-inducing trigger is when the lift rises from the ground to the elevated platform. This is triggered by button press by the participant. The lift takes about 18 seconds to travel to the platform. The simulated sensation of the lift rising is greatly enhanced by the accompanying audio. The ambient audio gradually changes from a loud construction scene to the sounds of traffic from the top of a building, while the sound of the lift's machinery plays in the foreground of the spatial audio zone. User testing showed that this greatly enhanced the feeling of rising in the lift.

Figure 8: A Fixable Part object emitting sparks after the Part Break event

3.4.2 Initial Part Break Event. This trigger was designed to set up the premise for the interaction task. After arriving at the platform, the user is given 5 seconds to get their bearings and look around. Following this short period, the machine parts located at various points on the platform will begin to malfunction with audio and visual cues. To prompt the user to exit the lift, an audio warning consisting of an automated voice, interspersed with an alarm bell and siren, will guide the user to step onto the platform, which is shown to be slightly raised (as with the haptic ledge). To enhance the jarring effect of the alarm, the controllers receive haptic signals, synchronized with the audio of the siren. A UI canvas also briefly informs the user of their task.

3.4.3 Lift Malfunction. After the user has been clear of the lift for 3 seconds, the lift will drop down and begin to malfunction. This involves the lift moving randomly between two bounded heights, at random time intervals, with accompanying sound and particle effects. From this point on, the user is unable to step back into the center of the platform, forcing them to engage with the threat of falling off the sides, which have no barriers. The behaviour of the lift malfunction is dependent on the state of the simulation, which is progressed either by successful task completion (fixing a part), or automatically based on an active timer. Please see Appendix A for a detailed description of lift malfunction behaviour.

3.4.4 Tool Loss. When triggered to lose the tool, a strong haptic buzz is sent to the controller gripping the tool and it is then forced from the user's grip. This causes the tool to either fall off the side of the platform, or down the central shaft, and clang on its way down, or fall to the ground on the platform and break apart. This is dependent on the velocity and direction that it is moving at, before the loss trigger. These two events were shown to be equally effective in alerting the participant to the loss of their tool in user testing.

4 EXPERIMENT DESIGN

To evaluate the research questions, two tests were designed.

(1) A within-subjects study to determine whether the VE would elicit fear in participants
(2) A between-subjects study to determine whether there is a significant difference in the degree of fear elicited in the restricted agency (Tool Loss) and control (No Loss) scenarios

Each study utilized both objective, psychophysiological measures of emotion, as well as subjective measures, which took the form of several self-report questionnaires (SRQ), evaluated post-immersion.

4.1 Participants

Participants were acquired using the UCT Psychology Department's ACSENT Lab email server. The dean of the Computer Science Department was also approached to distribute the advertisement to undergraduate Computer Science students. The resulting participant pool was made up of undergraduate Psychology and Computer Science students who responded to our request for participation and met our exclusion criteria. Each participant was given R50 remuneration for their time, in cash.

The exclusion criteria were developed to ensure that participants who may have been predisposed to distress were not harmed from VE exposure. As such, they included screening measures for general anxiety disorder, clinical acrophobia, as well as post-traumatic stress disorder.

To account for each of these, any advertisement respondents were required to score below 45 on the *State-Trait Anxiety Inventory* (STAI), below 45.45 on the *Acrophobia Questionnaire–Anxiety Subscale* (AQ-Anxiety), and below 3 on the *Primary Care PTSD Screen for DSM-5* (PC-PTSD-5). Please see Appendix A for a thorough explanation of all three exclusion questionnaires.

Of the 217 respondents to the study, 108 were randomly selected to screen for the Fear Environment. Of these, 66 passed the digital screening, and 36 of these participated in the study over the 9-day testing period. 1 participant opted to stop the simulation early, so their results were not included in the study, resulting in a final sample of 35 participants. See Appendix B for more detailed descriptive statistics about the participants.

4.2 Masking of Testing Purpose

Prior to immersion, the emotionally evocative nature of the environment was intentionally obscured from the participants. This was done in order to avoid the Hawthorne effect, which suggests that user behaviour changes when specific testing outcomes are known [4]. Premature disclosure of experiment purpose has been shown to produce skewed results in related SRQs [5], so was avoided by proposing the VE as a training simulation for construction workers in an elevated work scenario. The American Psychological Association's criteria for valid deception of research participants [6] were used as a guideline for the safe and justified masking of research. The deception involved in the study was justified by its benefit, the novel contribution that this research will make to current literature, given that there was no alternative method for completing the study.

Any potential for this deception to cause physical pain or severe emotional distress to participants, was mitigated by the strict exclusion criteria for participation, already outlined. Additionally, the nature of the deception was revealed to all participants during debriefing, which occurred as soon as the experiment was completed. This allowed users to understand the purpose of the research and presented them with a choice to have their results withheld, should they feel that the deception was unfair. Out of all 35 participants who completed the VE, none expressed any problems with the deception, so all their results were used.

4.3 Measures

4.3.1 Psychophysiological measures. Objective measures of emotion taken include heart rate (HR), respiration rate (RR) and skin conductivity level (SCL). Throughout the literature, positive changes in each of these measures have been associated with the sympathetic arousal of the autonomic nervous system [33]. In terms of emotional elicitation, this means that an involuntary fear response, generated as the body experiences some aversive trigger, can be associated with an increase in HR, RR and SC [32].

Psychophysiological measures were acquired with a Vrije Universiteit – Ambulatory Monitoring System (VU-AMS) device and analysed with the VU Data Analysis and Management Software (VU-DAMS) interface. To measure HR and RR the device used a configuration of 7 electrodes, placed around the front and back of the participant's torso. To measure SCL, measured in micro-Siemens (µS), the device used two nodes, which were attached to the middle and index finger of the user's weak hand with Velcro straps. An isotonic gel was applied to the transistors on the nodes to ensure a good reading from the device.

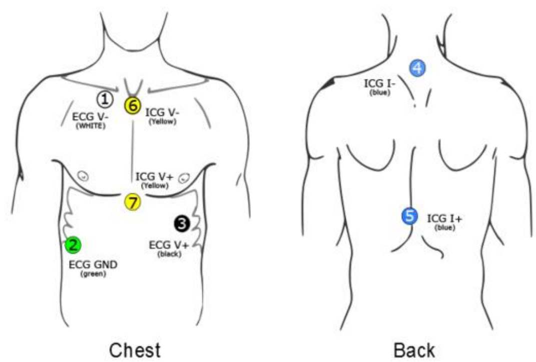

Figure 9: The electrode configuration for measurement of HR and RR with the VU-AMS device

The following subjective measures of emotion were also used:

4.3.2 The Visual Analogue Scale (VAS). This is a variation of Gross's measure of emotion elicitation in films [7, 8] which assesses how participants feel with a rating for 7 possible discrete emotions. In this case, however, the participant was asked to rate only the emotion that they felt most prominently during VE immersion. In using such a simple initial measure, we hoped to preserve the participant's overall impression of the VE, without giving them reason to endorse other emotions.

4.3.3 Discrete Emotions Questionnaire (DEQ). This instrument is designed to detect eight distinct emotions: anger, anxiety, desire, disgust, fear, happiness, relaxation, and sadness. The participant must give a rating to a collection of simple words used to describe emotions. It has been validated and compared against similar questionnaires, where it was found to be more sensitive at detecting self-reported emotions [9]. This is due to its ability to identify emotions that would otherwise be discounted by null effects on other measures of self-report [9].

4.3.4 Self-Assessment Manikin Scale (SAM). This is a 5-point assessment that captures the intensity of an emotion by visualising the elicited emotion's valence, arousal, and dominance [7,10]. Respectively, these characterise an emotion by the positive or negative feeling that it elicits, the strength of that feeling [7] and the participant's perceived ability to control the emotion in question [11]. This scale has been validated in previous VR research of emotion elicitation, which found similar dimensional ratings of emotion to real-world stimuli [7]. Please see Appendix D for a reference image of the SAM's 3 subscales.

4.4 Procedure

4.4.1 Screening. Participants who respond to the advertisement are asked to fill in digital versions of the screening questionnaires. After completion, their results are randomly split between the Fear environment, and a related Sadness environment. All respondents are sent response emails either sensitively indicating the reason that they were excluded from participation, with relevant contact details for support resources, or offering them an experiment timeslot, if they meet the criteria.

4.4.2 Pre-Experiment. The participant is given an informed consent form to read through and sign. They are then asked for demographic information regarding previous VR and general video game experience. With the participant's consent, the VU-AMS is attached to monitor HR, RR and SCL. Electrodes are applied by a lab assistant of the same gender that the participant reports as. *R*ecording is then started for resting baseline measures and a timestamp is taken.

4.4.2 VR Briefing and Guidance. After attaching VU-AMS the user is given a demonstration of controller use. They are also given general guidance about VR usage and interaction. Once baseline measures have been taken, and another timestamp is set, the participants are then fitted with the headset and then handed the controllers. They are given around 20 – 30 seconds to get used to the visuals and control scheme in the default SteamVR menu.

4.4.3 Tutorial. A timestamp is taken upon demo start. The participant then enters a tutorial, where they are shown the controls in better context and are briefed on what they can expect to be doing in the actual environment. At the end of the tutorial the participant presses a button to blend seamlessly between the tutorial and VE.

4.4.4 VE Immersion. A timestamp is taken when the VE loads in. The participant presses a button which causes the lift they are standing on to rise from floor level to an elevated platform over an 18 second period. The participant will then experience either the Tool Loss or No Loss treatment of the VE. In the Tool Loss treatment, an extra timestamp is taken at the point of tool loss. A final timestamp is taken when the environment ends.

4.4.5 Post-immersion. The user is then asked a series of SRQs. pertaining to the emotions that they experienced during VE immersion. Post-immersion physical measures are also taken in this period. The number of parts that the user was able to fix is noted down. Finally, the user is debriefed with regards to the masking of the experiment, signs a proof of payment form and is given R50 remuneration.

5 RESULTS AND DISCUSSION

5.1 Hypothesis 1 (H1): Post-immersion measures will indicate higher levels of psychophysiological arousal than baseline

The hypothesis predicted that the environment would produce a fear response in participants across the entire sample which would result in an increase in average HR, RR, and SCL. As such, a within-subjects study was utilized to compare mean, resting baseline and post-immersion measures for change in each trial. Both conditions were measured over a period of 1 minute 30 seconds. A post-immersion period was chosen to measure against baseline as it meant that users could be monitored at rest, rather than while moving around during VE task completion. These averages were compared using a paired one-tailed t test if they met parametric assumptions of normality, or a Wilcoxon Signed Rank test for paired samples if they did not. Throughout the testing period, 1 participant could not be measured with the monitoring device due to hardware failure, and a further 7 participants' data were found to be corrupt after recording. This resulted in a final sample of N=26.

5.1.1 Heart Rate. Baseline and post-immersion measures of HR both passed the Shapiro-Wilk test for normality ($p > 0.05$). Using a paired sample, one-tailed t-test it was determined that there was a statistically significant difference between baseline (M= 92,14, SD = 15,53) and post-immersion HR (M= 99,33, SD = 17,75), t(25) = -6,16, p<0,001, d = 1,21 (large effect size).

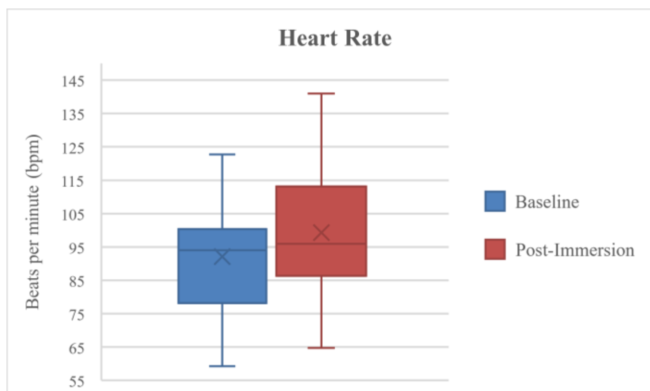

Figure 11: Box and whisker chart of HR

5.1.2 Respiration Rate and Skin Conductivity Level. The baseline and post-immersion measures for RR and SCL both failed the Shapiro-Wilk ($p > 0.05$). As such, they were each tested with a one-tailed Wilcoxon Signed-Rank test for paired samples. RR did not display a significant difference between baseline and post-immersion (Z-score = 0.81; p-value(exact) = 0.21; effect r = 0,81). SCL was shown to have a significant increase between baseline and post-immersion measures (Z-score = 2,84, p-value (exact) = 0.002, effect r = 0,42 (moderate effect size)). Due to the significance found in the measures of HR and SCL, the hypothesis was supported in most instances of psychophysiological markers of fear, however it was not supported in the case of RR.

Measure	Condition Mean(Std.Dev)		Z-score	Effect Size (r)	p-value (exact)
	Baseline	Post-Immersion			
RR	18,22(2,47)	17,29(3,14)	0,81	0,12	0,21
SCL	6,30(3,49)	7,35(4,42)	2,84	0,42	0,002

Table 1: Wilcoxon Signed-Rank test for paired samples, analysis for RR and SCL

5.2 Hypothesis 2 (H2): A triggered event which prevents the user from completing the environment's stated interaction task (Tool Loss) will cause a significant increase in these same psychophysiological markers, when compared to a control group (No Loss).

To test the hypothesized difference between the two treatments, a between-subjects study was designed. The test would take measurements from the last 40 seconds of VE immersion in both treatments and subtract the average measures from the preceding period of VE immersion, approximately 4 minutes 20, in order to gauge whether there is a significant difference between the change in psychophysiological measures between conditions.

The datasets created from this passed the Shapiro-Wilk (p > 0,05) for all 3 psychophysiological measures, so one-tailed independent-groups t tests were used to compare the change values between conditions.

The agency-restricted Tool Loss group (M = 4,68, SD = 5,1) had a significantly greater change in average HR than the No Loss group (M = 0,56, SD = 4,43), t(24) = 2,19; p = 0,02; d = 0,85 (large effect size). The Tool Loss group did not show a statistically significant increase in either SCL or RR over this period, however. The results are summarized in the table below.

Measure	Condition Mean(Std.Dev)		df	t(df)	p-value	Cohen's d
	No Loss	Tool Loss				
HR	0,56(4,43)	4,68(5,1)	24	2,19	0,02	0,851501
RR	1,14(3,32)	-0,98(3,53)	22	1,5	0,07	0,62
SCL	0,006(0,75)	0,22(0,63)	21	0,79	0,22	0,33

Table 2: Independent-groups one-tailed t tests for H2

In this case, the only measure that supported this hypothesis was HR.

To identify if this external factors, the mean AQ-Anxiety scores, as well as the mean VR and Video Game Experience scores, of the two testing groups were analysed. None of the three datasets passed Shapiro-Wilk tests (p > 0,05) so each was tested with a Mann-Whitney Test for two independent samples. Neither Experience score was shown to have a significant difference between the Tool Loss Conditions. The results for the one-tailed test on the AQ-Anxiety scores, however, showed that the scores for participants in the Tool Loss condition (M = 26,76; SD = 10,86; Min = 3, Max = 45) were significantly higher than the scores of the No Loss condition (M = 15,05; SD = 13,88; Min = 0; Max = 42), Z = 2,59; p-value (exact) = 0,004; r = 0,44 (moderate effect size). The mean AQ-Anxiety scores of both groups (TL = 26.76, NL = 15,06) lie further than 2 standard deviations from the mean score of both a significant low acrophobia sample (M= 9,87; SD = 3,85), as well as a high acrophobia sample (M = 60,54; SD = 13,95), as validated in the literature [27]. This suggests that, despite the significant difference between the samples, both datasets lie within a set of moderately acrophobic scores.

5.3 Hypothesis 3 (H3): Subjective measures of emotion will indicate that, during immersion, users experienced fear significantly more than other emotions. This will be enhanced in the Tool Loss treatment.

5.3.1 Visual Analogue Scale. The answers to the VAS question were tallied and compiled as the frequency table below.

	No Tool Loss			Tool Loss		
Emotion	**Count**	**Mean**	**Std Dev**	**Count**	**Mean**	**Std Dev**
Fear	4	7	1,29	3	6,33	1,53
Neutral	2	5,5	0.71	0	0	0
Amusement	10	8,9	1,36	9	8,33	1,32
Surprise	2	8,5	0,71	5	8,8	1
Sadness	0	0	0	0	0	0
Contentment	0	0	0	0	0	0
Disgust	0	0	0	0	0	0

Table 3: Frequency Table of the VAS results for primary emotion felt, mean scores taken from a rating (1-10)

As can be seen from the table, Fear is the primary emotion endorsed in only 22,22% of trials in the No Loss treatment, and 17,65% of the Tool Loss treatment. Amusement was more commonly chosen as the primary emotion experienced during both treatments, comprising 56% of all No Loss trials, and 52,94% of all Tool Loss trials. When endorsed, on average, Amusement also had a higher rating of severity from the participants in both treatments.

5.3.2 Self-Assessment Mannequin None of the subscales for the SAM were shown to have normal distribution according to Shapiro-Wilk (p > 0.05) testing. As such, each was compared between conditions using a Mann-Whitney Test for two independent samples.

A test comparing valence did not show a significant difference in the scores between the Tool Loss (M = 1,94; SD = 0,97) and No Loss (M = 1,78; SD = 0,88) conditions, Z = 0,48; p-value (exact = 0,33) r = 0,08).

A test comparing arousal showed that there was a significant difference between the Tool Loss (M = 3,11; SD = 0,3) and No Loss (M = 1,94; SD = 0,8) environments, Z = 2,77; p-value (exact) = 0,003; r = 0,46 (moderate effect size). This indicates that the average score for arousal in the Tool Loss treatment was significantly higher than the average score for the No Loss treatment.

A test comparing dominance showed that there was a significant difference between the Tool Loss (M = 3,29; SD = 1,31) and No Loss (M = 4,28; SD = 0,89) conditions, Z = 2,27; p-value (exact) = 0,01; r = 0,38 (moderate effect size). This shows that the average score for dominance in the Tool Loss treatment was significantly lower than the average score for the No Loss treatment.

5.3.3 Discrete Emotions Questionnaire. None of the subscale datasets for the DEQ passed the Shapiro-Wilk ($p > 0.05$) for both treatments, so Mann-Whitney tests for Two Independent Samples were utilized to compare the conditions. Of the 8 emotion subscales, there was no significant difference between the two conditions in any case. This is outlined in the table below.

Emotion	Treatment	Mean	Std. Dev	Mann-Whitney U	Z-score	Effect Size (r)	p-value (exact)
Anger	No Loss	1,23	0,5	149	0,14	0,02	0,91
	Tool Loss	1,38	0,82				
Disgust	No Loss	1,25	0,44	137,5	0,54	0,09	0,61
	Tool Loss	1,39	0,53				
Fear	No Loss	2,54	1,69	141	0,38	0,06	0,71
	Tool Loss	2,75	1,75				
Anxiety	No Loss	2,4	1,32	135,5	0,56	0,09	0,57
	Tool Loss	2,82	1,8				
Sadness	No Loss	1,44	0,99	151	0,05	0,01	0,96
	Tool Loss	1,46	0,83				
Desire	No Loss	2,4	1,24	110	1,4	0,24	0,16
	Tool Loss	3,29	1,76				
Relaxation	No Loss	3,74	1,81	139,5	0,43	0,073	0,66
	Tool Loss	4,03	1,97				
Happiness	No Loss	5,21	1,59	143,5	0,3	0,05	0,76
	Tool Loss	4,88	1,86				

Table 4: Mann-Whitney Test for Two Independent Samples, analysis of DEQ results

Further analysis was done to compare the mean score for the Fear subscale to other emotional subscales. Using a one-tailed Wilcoxon Signed-Rank test, it was determined that mean scores on the Happiness and Relaxation subscales were significantly higher than the Fear subscale in both conditions (NL: Z-score = 22,96; p-value(exact) < 0,001; effect r = 0,49 (moderate effect size)), (TL: Z-score = 2,33; p-value(exact) = 0,008; effect r = 0,4 (moderate effect size)).

5.4 Discussion

In the assessment of H1, HR and SCL were both found to increase significantly between baseline and post-immersion measures. However, due to RR not showing a significant difference between conditions, H1 could not be said to be fully supported.

A potential explanation for the lack of the expected, significant increase in RR between conditions is the cognitive nature of the interaction task in the VE. Research on the effects of cognitive load on RR has shown that the anticipation of a cognitive task, alone, can be enough to raise pre-task RR above reliable baseline measures [14]. It is possible then that baseline measures of RR could have been affected by anticipation of entering the VE, which participants knew would require them to focus on a mentally demanding construction task. Considering this, we could refine H1 to apply to the measurement of HR and SCL only. In this amended case, we can conclude that H1 was supported, indicating that there was significant evidence of psychophysiological changes arising from immersion in the VE. Furthermore, these changes were found to be consistent with sympathetic arousal of the autonomic nervous system, as a result of a fear response.

When testing for significant change in psychophysiological measures between Tool Loss and No Loss treatments during the last 40s of the VE (H2), only HR showed a statistically significant difference between conditions. The analysis showed that HR during the period where user agency was restricted in the Tool Loss scenario increased, on average, significantly more than the same period in the No Loss scenario.

The lack of significant difference in change of SCL could be explained by the fundamental method that SCL uses to measure skin conductance. While phasic measures of skin conductance, such as Skin Conductance Response (SCR), can be used to identify changes in skin conductance between discrete events and triggers, SCL reflects a tonic, or sustained, change in sympathetic activation of the autonomic nervous system [23]. As such, the 40 second period measured was not sufficiently long enough for significant changes in SCL to occur. Unfortunately, SCR requires extended periods of rest between triggers in order to provide valid results [24]. These added measures could not have been reasonably added to the experiment procedure, given time constraints, as well as the script of the VE.

Similarly to H1, it is likely that the cognitive nature of the interaction task in the environment can explain the lack in significant difference in the change of RR between conditions. Research suggests that respiratory patterns are strongly influenced by unique combinations of individual personality traits, as well as task characteristics [14]. For example, an appraisal of a fear-inducing task between similarly phobic participants could lead to coping responses that vary between slow, strained breathing, characterized by low RR, to intense hyperventilation, which would indicate increased RR [14]. To account for these factors, any measures of RR would need to be accompanied by complex analysis of other demographic information. Due to time constraints, it was not possible to carry out any further analysis of the RR data in this case.

In the absence of reliable measures of RR and SCL, H2 was not conclusively supported by the data. Despite this, the increase in HR in the Tool Loss scenario is still of interest, as it presents a significant marker of the arousal of the autonomic nervous system as a result of an agency-restricting event.

Further analysis of the demographic data pertaining to the two groups revealed that the AQ-Anxiety scores of the Tool Loss group were significantly higher than the No Loss group. This further complicates the validity of H2, as it suggests that the Tool Loss group's predisposition to greater acrophobia than the No Loss group could have influenced the increased psychophysiological response, in addition to the agency restricting event. This is unlikely to be the case, however, as both groups displayed healthy, moderately phobic mean AQ-Anxiety scores, when compared to previously validated samples.

From the analysis of the DEQ and VAS we cannot find any statistically significant indication that fear was reported as the primary emotion experienced during VE immersion. In fact, more than half of the entire participant sample reported Amusement as their primary emotion experienced on the VAS, while analyses of the DEQ results revealed that participants scored subsets of words pertaining to Happiness and Relaxation significantly higher than the subset related to Fear.

Though experience of excited amusement is consistent with the results found in H1, a high score for relaxation is at odds with the psychophysiological measures of arousal found. Furthermore, this fails to explain the increase in HR found between the conditions in H2. A potential explanation for this discrepancy can be found in the demographic data of the participant pool. The participant sample was shown to have little previous VR experience on average, as well as being comprised of an age range who are likely to be interested in technology. Because of this, the novel experience with VR technology was likely to be a positive one, regardless of the contents of the simulation. This could have caused users to report their positive feelings of VR use rather than their response to the emotion-inducing content of the VE.
In addition, as the sample was screened in advance to remove any phobic participants, the content of the simulation, while dangerous, may not have been overtly frightening to participants. This may have had the effect of inducing excitement and pleasure in participants, such as in the case of extreme aerial sports, like skydiving or bungee jumping.

Results from the VAS help to support this theory. Comparisons of valence scores showed that there was no significant change in non-verbal reported happiness between the Tool Loss and No Loss treatments. Despite this, scores for arousal and dominance showed that participants in the Tool Loss scenario reported a significantly higher feeling of general excitement and a significantly lower feeling of control over their emotions during immersion. These results are consistent with the findings of H1 and H2, indicating sympathetic arousal and a loss of agency. The VAS findings help to reveal the shortcomings of verbal measures of emotion in this case, as they fail to explain the difference in HR measured between the conditions.

6 CONCLUSIONS

The aim of the first research question was to determine whether the VE, developed specifically for this study, would be effective in inducing a fear response in a non-phobic sample. Results from H1 showed significant increases in measures of HR and SCL between baseline and post-immersion. This could be indicative of arousal due to sympathetic activation of the autonomic nervous system. As such, the VE can be said to have induced psychophysiological changes analogous to a fear response. Subjective measures of emotion were not successful in identifying fear as the primary emotion experienced by participants, however.

The aim of the second research question was to determine whether an event that limits user control, in an already fear-inducing VE, will lead to heightened measures of fear. HR was determined to be the only valid psychophysiological measure of fear in H2. Results showed that it increased significantly more in the Tool Loss condition, from average preceding immersion measures, than the control condition. This indicates that a loss of user control may have resulted in a higher state of fear arousal. Non-verbal, subjective, self-report measures of general excitement (arousal) and sense of control (dominance) indicate similar subjective emotional response. From this, there is strong evidence to suggest that the variable of user control was influential in creating a heightened fear response between treatments.

7 LIMITATIONS AND FUTURE WORK

One major limitation in the evaluation of the subjective measures of emotion was the lack of baseline measures, taken before immersion. Though results could still be acquired in the between-subjects study, it limited the application of the within-subjects subjective study as emotional effects that arose prior to VE immersion could not be accounted for.
Another limitation is the size of the sample. Unfortunately, due to strict testing schedule and a shared work environment, there was limited time to carry out the experiment. Since my project partner and I could not share a sample, we had to divide up the respondents further, and this led to the lower sample size that we dealt with. A higher sample size would have led to a greater chance of identifying potential trends in statistical analysis, as well as providing greater validity for trends established in this paper. Additionally, if a sample with little VR experience is used, the use of subjective measures of emotion should be reconsidered, as participants may endorse feelings relating to their novel experience of the technology, rather than the content of the VE.

The strict timescale also had the effect of reducing the
Methodological improvements for the measurement of the agency-restricting trigger event could include the use of a phasic measure of skin conductance, rather than SCL. Additionally, the length of the periods that the various physical measures were taken over, as well as the overall length of VE immersion for each participant was quite restricted, given time constraints. Longer periods of measurement are likely to have shown more accurate psychophysiological readings.
Future studies with a larger participant pool could aim to add further experimental conditions so that the effects of agency restriction can be tested in better depth. Potential demographics of interest could include: a phobic vs non-phobic group, a high VR experience vs low VR experience group, or even an environment with no user interaction to test as a control.

8 REFERENCES

[1] Philip Lindner et al. 2017. Creating state of the art, next-generation virtual reality exposure therapies for anxiety disorders using consumer hardware platforms: Design considerations and Future Directions. Cognitive Behaviour Therapy 46, 5 (2017), 404–420. DOI:http://dx.doi.org/10.1080/16506073.2017.1280843

[2] Michael Meehan, Sharif Razzaque, Brent Insko, Mary Whitton, and Frederick P. Brooks. 2005. Review of four studies on the use of physiological reaction as a measure of presence in stressful virtual environments. Applied Psychophysiology and Biofeedback 30, 3 (2005), 239-258. DOI: https://doi.org/10.1007/s10484-005-6381-3

[3] Julia Diemer, Lohkamp, N., Muhlberge., A & Zwanzger P. 2016. Fear and physiological arousal during a virtual height challenge—effects in patients with acrophobia and healthy controls. Journal of Anxiety Disorders. 37:30-39. DOI: https://doi.org/10.1016/j.janxdis.2015.10.007

[4] Berkely Rice. 1982. The Hawthorne defect: Persistence of a flawed theory. Psychology Today, 16(2), 70-74

[5] Daniel Västfjäll. 2001. Emotion induction through music: A review of the musical mood induction procedure. Musicae Scientiae 5, 1_suppl (2001), 173–211. DOI:http://dx.doi.org/10.1177/10298649020050s107

[6] American Psychological Association (APA). 2002. Ethical principles of psychologists and code of conduct. Available at: http://www.apa.org/ethics/code/index.aspx.

[7] Giuseppe Riva, Fabrizia Mantovani, Claret Samantha Capideville, Alessandra Preziosa, Francesca Morganti, Daniela Villani, Andrea Gaggioli, Cristina Botella, and Mariano Alcañiz. 2007. Affective interactions using virtual reality: the link between presence and emotions. Cyberpsychology & Behavior 10, 1 (Feb 2007), 45-56. DOI: https://doi.org/10.1089/cpb.2006.9993

[8] James J. Gross and Robert W. Levenson. 1995. Emotion elicitation using films. Cognition & Emotion 9, (1995), 87–108.

[9] Cindy Harmon-Jones, Brock Bastian, and Eddie Harmon-Jones. 2016. The Discrete Emotions Questionnaire: A new tool for measuring state self-reported emotions. PLOS ONE 11, 8 (2016). DOI:http://dx.doi.org/10.1371/journal.pone.0159915

[10] Cindy Harmon-Jones, Brock Bastian, and Eddie Harmon-Jones. 2016. The Discrete Emotions Questionnaire: A new tool for measuring state self-reported emotions. PLOS ONE 11, 8 (2016). DOI:http://dx.doi.org/10.1371/journal.pone.0159915

[11] Rosalind W. Picard, Elias Vyzas, and Jennifer Healey. 2001. Toward machine emotional intelligence: Analysis of affective physiological state. IEEE Transactions on Pattern Analysis and Machine Intelligence 23, 10 (2001): 1175-1191. DOI:10.1109/34.954607

[12] Mel Slater. 2009. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364, 1535 (Dec 2009), 3549-3557. DOI: https://doi.org/10.1098/rstb.2009.0138

[13] Daniel Gromer, Octávia Madeira, Philipp Gast, Markus Nehfischer, Michael Jost, Mathias Müller, Andreas Mühlberger, and Paul Pauli. 2018. Height simulation in a virtual reality CAVE system: validity of fear responses and effects of an immersion manipulation. *Frontiers in Human Neuroscience* (Sep 2018): 372. DOI: https://doi.org/10.3389/fnhum.2018.00372

[14] Mariel Glassmann, Elke Vlemincx, Andreas Von Leupoldt, Justin M. Mittelstädt, and Omer Van den Bergh. 2016. Respiratory changes in response to cognitive load: a systematic review. *Neural plasticity*, (June 2016), 8146809. DOI:https://doi.org/10.1155/2016/8146809Andreas

[15] Joanna Kisker, Thomas Gruber, and Benjamin Schöne. 2021. Behavioral realism and lifelike psychophysiological responses in virtual reality by the example of a height exposure. *Psychological research* 85, 1 (Feb 2021), 68-81. DOI: https://doi.org/10.1007/s00426-019-01244-9

[16] Alessandra Gorini, Claret S. Capideville, Gianluca De Leo, Fabrizia Mantovani, and Giuseppe Riva. 2011. The role of immersion and narrative in mediated presence: the virtual hospital experience. *Cyberpsychology, Behavior, and Social Networking* 14, 3 (Mar 2011), 99-105. DOI: https://doi.org/10.1089/cyber.2010.0100

[17] Kyle E. Madsen. 2016. The differential effects of agency on fear induction using a horror-themed video game. *Computers in Human Behavior* 56, (Mar 2016), 142-146. DOI: https://doi.org/10.1016/j.chb.2015.11.041

[18] Margaret M. Bradley, and Peter J. Lang. 2007. Emotion and motivation. In John T. Cacioppo, Louis G. Tassinary, and Gary Berntson, eds. Handbook of psychophysiology. 2007. Cambridge University Press DOI: https://doi.org/10.1017/CBO9780511546396.025

[19] Jih-Hsuan Tammy Lin. 2017. Fear in virtual reality (VR): Fear elements, coping reactions, immediate and next-day fright responses toward a survival horror zombie virtual reality game. *Computers in Human Behavior* 72, (2017), 350-361. DOI: https://doi.org/10.1016/j.chb.2017.02.057

[20] Jih-Hsuan Tammy Lin, Dai-Yun Wu, and Chen-Chao Tao. 2018 So scary, yet so fun: The role of self-efficacy in enjoyment of a virtual reality horror game. *New Media & Society* 20, 9 (Sep 2018), 3223-3242. DOI: https://doi.org/10.1177/1461444817744850

[21] Edna B. Foa, and Michael J. Kozak. 1986. Emotional processing of fear: exposure to corrective information. Psychological bulletin 99, 1 (Jan 1986), 20-35. DOI: https://doi.org/10.1037/0033-2909.99.1.20

[22] Mel Slater. 2018. Immersion and the illusion of presence in virtual reality. The British Journal of Psychology, 109(3), 431–433. DOI: https://doi.org/10.1111/bjop.12305

[23] Michael E. Dawson, Anne M.Schell, and Diane L. Filion. 2007. The Electrodermal System, In John T. Cacioppo, Louis G. Tassinary, and Gary Berntson, eds. Handbook of psychophysiology. 3rd edn. Cambridge: Cambridge University Press, pp. 159–181. DOI:https://doi.org/10.1017/CBO9780511546396

[24] Vrije Universiteit Amsterdam. 2022. *Data Analysis and Management Software (Dams) For The Vrije Universiteit Ambulatory Monitoring System (Vu-Ams).* (January 2022). Retrieved September 14, 2022 from http://www.vu-ams.nl/support/instruction-manual/ [15/09/22]

[25] Bruce L. Baker, David C. Cohen, and Jon Terry Saunders. 1973. Self-directed desensitization for acrophobia. *Behaviour Research and Therapy* 11, 1 (1973): 79-89. DOI: https://doi.org/10.1016/0005-7967(73)90071-5

[26] Bruce Arroll, Henry B. Wallace, Vicki Mount, Stephen P. Humm, and Douglas W. Kingsford. 2017. A systematic review and meta-analysis of treatments for acrophobia. *Medical Journal of Australia* 206, 6 (2017): 263-267. DOI: https://doi.org/10.5694/mja16.00540

[27] Shari A Steinman and Bethany A Teachman. 2011. Cognitive processing and acrophobia: validating the Heights Interpretation Questionnaire. *Journal of Anxiety Disorders* 25,7 (2011), 896-902. DOI:10.1016/j.janxdis.2011.05.001

[28] Charles D. Spielberger. 1989. *State-trait anxiety inventory*: Bibliography (2nd ed.). Palo Alto: Consulting Psychologists Press.

[29] Gina Gilpin, James Gain, and Gosia Lipinska. 2021. The physiological signature of sadness: A comparison between text, film and virtual reality. *Brain and Cognition* 152 (2021), 105734. DOI:http://dx.doi.org/10.1016/j.bandc.2021.105734

[30] Ozcan Kayikcioglu, Sinan Bilgin, Goktug Seymenoglu, and Artuner Deveci. 2017. State and trait anxiety scores of patients receiving intravitreal injections. *Biomedicine Hub* 2, 2 (2017): 1-5. DOI:10.1159/000478993

[31] International Society for Traumatic Stress Studies (ISTSS). 2015. *Primary Care PTSD Screen for DSM-5 (PC-PTSD-5).* Retrieved July 26, 2022 from https://istss.org/clinical-resources/assessing-trauma/primary-care-ptsd-screen-for-dsm-5-(pc-ptsd-5)

[32] Sylvia D. Kreibig. 2010. Autonomic nervous system activity in emotion: A review. *Biological psychology* 84, 3 (July 2010), 394-421. DOI:https://doi.org/10.1016/j.biopsycho.2010.03.010

[33] Erel Shvil, Heather L. Rusch, Gregory M. Sullivan, and Yuval Neria. 2013. Neural, psychophysiological, and behavioral markers of fear processing in PTSD: a review of the literature. *Current psychiatry reports* 15, 5 (May 2013), 1-10. DOI: https://doi.org/10.1007/s11920-013-0358-3

Appendix A

No. Parts Fixed/States Progressed	Maximum Length	Lift Behavior		
		Movement	SFX	Particle Effects
0	2 min 20 sec	At baseline malfunction behaviour, the lift will move to a new random y-axis location every 4-6 seconds.	Upon moving, the lift will play the sound of a rolling mechanism, chosen at random from a list, and then a heavy metal clang when coming to rest	The lift has sparking particle emission at the point where it meets the runner mechanism of the support shaft and at each of its 4 corners.
1	1 min 40 sec	Lift movement will become more erratic, decreasing the randomized interval between movements to 2-4 seconds.	The previous sound effects are altered to include higher-pitched sounds of stressed metal and louder mechanical groaning	Same as above
2	45 seconds	The lift will stop moving at this point, regardless of its final movement endpoint	The groaning metal sounds worsen and further sounds indicating a stressed winch mechanism/cable are added to the surround audio.	Particle emission is halted while the lift is stationery
3	15 seconds	The lift will remain stationary for another 15 seconds, and then will drop back to the street level, below	The groaning and stressed metal sounds worsen until the 15 seconds are over. After this point, a loud bang is played, and a rapidly turning winch sound plays, to indicate that the lift is in freefall.	Particle emission continues once the lift starts to fall

Table A1: Detailed overview of the malfunction behaviour of the lift over the course of the VE

Appendix B

Descriptive Statistics	***Tool Loss (n = 17)*** ***Mean (Std. Dev)***	***No Loss (n = 18)*** ***Mean (Std. Dev)***
Gender (%F)	53%	61%
No. Parts Fixed	1.18 (0.73)	1.39 (0.92)
Prior VR Experience	0.41 (0.62)	0.39 (0.70)
Prior Video Game Experience	2.94 (1.64)	2.94 (1.59)
AQ-Anxiety	26.76 (10.86)	15.06 (13.88)
Presence Rating (1-10)	7.18 (2.19)	7.33 (1.78)

Table B1: Descriptive Data for Participants

Rating	**Condition**
0	No experience
1	Minimal video game experience, unfamiliar with game controller
2	Limited video game experience, familiar with game controller
3	Moderate video game experience
4	Very experienced, possibly in the past (not a "gamer")
5	Own and regularly use game console/pc for gaming

Table B2: Categories for scoring prior video game experience

Rating	**Condition**
0	No experience
1	Minimal experience (1 <= x < 3 times)
2	Limited experience (3 <= x < 5 times)
3	Moderate experience (5 <= x < 10 times)
4	Very experienced
5	Own and regularly use VR

Table B3: Categories for scoring prior VR experience

Appendix C

Acrophobia Questionnaire – Anxiety Subscale (AQ-Anxiety)
The AQ-Anxiety is a self-report questionnaire for assessing height-based [25] It is widely used in testing for clinical acrophobia, and has been shown to have replicable results [26]. Participants must rate their anxiety when faced with a situation on a 7-point scale (0=not anxious at all, 6=extremely anxious). Any participants who scored sufficiently high (≥ 45.45, as determined by Steinman and Teachman [2011] to be within one standard deviation below a subset of extreme acrophobics) were excluded from the experiment.

State-Trait Anxiety Inventory (STAI)
Finally, the STAI, a 20-question self-report questionnaire, will be used to determine whether participants suffer from general symptoms of anxiety, measured on a 4-point scale [28]. It has seen use as both a testing measure [7] and an exclusion criterion [29] in previous research involving the elicitation of emotion in VR, so was considered likely to be suitable for our study. Participants with an STAI score greater than 45 are classified as having high general anxiety [30] and were excluded from the experiment.

Primary Care PTSD Screen for DSM-5 (PC-PTSD-5)
This screening tool first asks participants whether they have ever experienced a significantly traumatic incident in their life, with some examples given. If they answer positively, they must answer a series of 5 binary, "yes" or "no", questions regarding the incident, aimed to identify symptoms of PTSD. Participants who answer "yes" to 3 or more of the questions are categorized as likely to be suffering from PTSD and were excluded from the experiment [31].

Appendix D

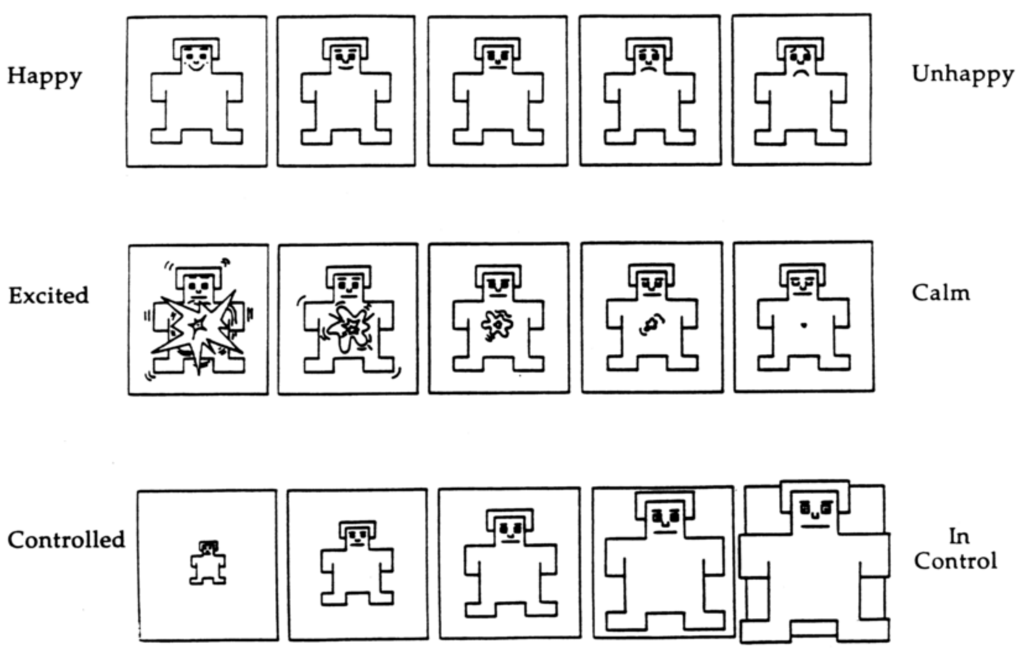

Figure D1: The SAM Scale with ratings for valence, arousal, and dominance (in descending order)