UNIVERSITY OF CAPE TOWN

DEPARTMENT OF COMPUTER SCIENCE

CS/IT Honours Project Final Paper 2022

Title: The Effect of User Interaction on Eliciting Sadness in Virtual Reality

Author: Julia Kinghorn

Project Abbreviation: VRSad

Supervisor(s): James Gain, 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 Interaction on Eliciting Sadness in Virtual Reality

Julia Kinghorn
Computer Science
University of Cape Town
City State Country
kngjul010@myuct.ac.za

ABSTRACT

This user study (n=43) tests how the elicitation of sadness is affected by different levels of user interaction in a Virtual Reality environment, with the aim of contributing to research on the use of Virtual Reality for the treatment of psychological disorders like post-traumatic stress disorder. The environment involves a narrative of loss, in which each user adopts a dog, plays with it, and then unexpectedly loses it when it is hit by a car. The environment has two levels of interactivity, one in which the dog and environment do not react to user input, so there is a low sense of interaction, and one in which the dog and environment are highly reactive, leading to a greater sense of interaction. Descriptive results indicate that altering user interaction levels can alter the type of sadness produced. It appears that higher levels of interaction were associated with an activating pattern of sadness, while lower levels induce a more traditionally recognised deactivating pattern of sadness.

KEYWORDS

Virtual Reality, Virtual Environment, Interaction, Sadness, Elicitation, Presence, Psychophysiological

1 Introduction

Virtual Reality (VR) involves immersing users in a 3D computer-generated environment that allows for real-time interaction with synthetic stimuli [2]. VR is used in psychology as a potential medium for Exposure Therapy (VRET). VRET desensitises patients with post-traumatic stress disorder (PTSD) and phobias by exposing them to the stimuli that trigger an extreme emotional response (triggers) in a Virtual Environment (VE) [3]. VR has the advantage of transporting patients beyond spatial and resource constraints, and enables patients to interact with their emotions in a safe space [4, 5, 6]. However, there is little baseline research on how best to elicit emotion in VR, which is important because of the heightened emotional response psychological triggers cause. Human experience of sadness is common, especially in psychological disorders including PTSD, phobias, depression, and anxiety, and often dictates how one behaves towards oneself and others [7]. Despite this, our understanding of it is limited. As such, the ability to elicit and study sadness in a controlled environment would be invaluable in the development of psychological treatments and knowledge. This study aims to contribute to baseline research on the feasibility of using VR in psychological applications.

Until now, research on eliciting emotion in VR has focused largely on the relationship between presence and emotion elicitation, or on the efficacy of VR versus other stimuli to elicit a range of emotions, as discussed below. This means that there is little literature on the best way to elicit sadness in VR. An important factor when it comes to the design of a VE is how interactive it is [38]. That is, how much the user can influence what happens in the VE, and how much the VE responds to user input. This study focuses on the effect of this element of VE design, to investigate whether a difference in interactivity affects the levels of sadness elicited by a VE. Essentially, this project aims to answer the question: "How does manipulating the amount of interaction with a dog in a VE affect the sadness elicited when the dog dies?". The VE developed to test this question is built from a previous iteration of the same environment.

2 Related Work

Although literature on the elicitation of sadness in VR is limited, there has been research on general emotion elicitation in VR, particularly in relation to presence. This means that while there is an adequate reference pool for this type of study, further considerations are necessary when tailoring the study to one about sadness and interaction.

2.1. Considerations for eliciting and measuring sadness

2.1.1. How to elicit sadness

Literature investigating a reliable method of eliciting sadness in a controlled setting is necessary but lacking [10]. It has been determined that emotion elicitation does not have a "one size fits all" medium [7], but that on average, visual stimuli, music, and autobiographical recall are most effective at inducing emotional response [7]. It is also suggested that testing a combination of these mediums may be more effective than testing them in isolation [7]. VR allows for a combination of mediums with the added power of presence, offering much potential as an effective method of elicitation [55].

Another significant factor in the success of an elicitation method is its content. It has been argued that a combination of certain types of situations, appraisal, and attribution must be experienced to elicit sadness [8]. This combination requires one to appraise a situation as some kind of loss, the cause of which cannot be attributed to anything. If one feels that someone else is responsible for the loss, one is likely to experience anger. However, when one feels that the situation is beyond anyone's control, sadness is likely to be experienced [8, 12]. The most affective versions of loss are related to the end of relationships (often by death) [8, 9, 11, 55].

However, it is important to note that this content is determined in a western context of psychology [1]. The cultures of a population must also be considered when determining emotions elicited by certain situations [8, 61], as culture determines how people feel it is appropriate to respond to certain content [8, 12].

The literature thus suggests that, to effectively elicit sadness, one should combine stimuli that create a sense of (culturally appropriate) loss, in such a way that the participant cannot perceive anyone as directly to blame. Personally relevant experiences of loss are particularly evocative [13, 14]. It is widely acknowledged that these experiences should come in a narrative form, as participants are more deeply immersed in an experience when they have aims to follow [14, 15, 16, 58, 59]. This is best achieved with short stories, as they allow participants to form personal and emotional connections to the narrative [17]. As such, this study hypothesises that a short narrative experience in which one gets a dog, bonds with it, and then loses it will elicit sadness.

2.1.2. How to measure and interpret sadness in VR

The collection of physiological data to identify sadness is evident in many studies, and has historically been linked to one combination of physiological responses [46]. However, according to Kreibig's [2010] meta-analysis of psychophysiological measures, it is likely that subtypes of sadness exist. These include 'deactivating' sadness, in which parasympathetic activation and a decreased respiration rate (RR) and heart rate (HR) is observed [11], and 'activating' sadness, in which parasympathetic withdrawal and an increased RR and HR is observed [10, 20, 46]. According to the literature, deactivating sadness is more likely if an inevitable or historical loss has occurred. Activating sadness is related to an imminent, but not inevitable, loss. This is usually found in situations where an individual has agency over a situation, and may view the loss as more of a failure than an occurrence by chance [55]. This suggests that either the literature on sadness inducing content reviewed above is too limited in its concept of loss without blame, or that activating sadness is a form of sadness combined with another emotion as a result of one's sense of accountability, such as guilt. More research is needed to clarify this, so this paper assumes that different types of sadness can be experienced, with a sense of accountability and without.

Rivu et al. [2021] found that VR-elicited sadness has no statistically significant difference from otherwise-elicited sadness. However, levels of arousal are slightly higher, and dominance slightly lower, in VR-elicited sadness. Gilpin et al.'s [2021] findings also suggest that the physiological markers of sadness change depending on the elicitation method used. The type of sadness they identified in VR was an activating pattern of sadness, which contrasted the deactivating sadness pattern they found that was caused by film clips with the same content as the VR [55]. This suggests that participants have a greater sense of agency in VEs, so the sadness induced by a VE feels more imminent. Furthermore, a participant's sense of agency may determine the sadness subtype elicited. In other words, greater agency leads to an activating pattern of response. These results are supported by Kreibig et al.'s [2007] study to test the response to sadness-inducing film clips, in which they found that the resulting sadness was deactivating. However, only eight participants were involved in this study, so the results are not necessarily generalisable [8].

Self-report measures of emotion are also used extensively in the literature. Broadly, emotions can be categorised according to several dimensions [46]. The most notable of these are valence, which defines an emotion as positive or negative, arousal, which indicates the strength of an emotion [4], and one's perception of control (e.g. [4, 18, 19, 46, 47]). Sadness is traditionally associated with low arousal and negative valence [4]. However, through the physiological methods discussed above, Gilpin et al. [2021] have found that the arousal level seems to change depending on the elicitation technique employed [55]. This means that sadness could be low or high arousal, depending on the context, and that discrete qualitative measures of sadness may not be as useful as other literature suggests. However, discrete measurements may be easier to obtain than dimensional ones, because some behaviours, subjective feelings, eliciting situations, and appraisals cannot be measured dimensionally [54].

2.2 How to make a VE to successfully elicit sadness

As it has been proven that VR can elicit sadness, the next step is to understand how best to use VR for this purpose, so that this study can ensure that all other variables are controlled in its attempt to isolate the effect of interaction. To create a VE that will effectively elicit emotion, there are a number of factors to consider beyond its content.

2.2.1 Presence:

Most literature emphasises immersion and presence. This paper defines immersion as what technology delivers in terms of sensory displays – the greater the number of sensory displays, the greater the immersion. Presence is determined by one's reaction to immersion, and describes the extent to which one feels that they are in a particular environment [14, 22, 23, 24]. Different levels of presence are reported by different people in the same immersive VE [25]. However, it is broadly reported that increased immersion results in increased presence [13, 14, 26, 27, 28, 29, 58].

There is debate around the relationship between immersion, presence, and emotion elicitation, but many regard presence as a necessity for eliciting emotions in VR [14, 16, 29, 30, 31, 32, 33, 34]. The correlation between emotion and immersion (thus presence) has been widely documented in the literature, especially in work about VRET [3, 14, 15, 16, 26, 27, 32, 33]. However, some report that presence and emotion have a non-linear relationship [13, 14, 35]. Studies which report a linear relationship focus on strongly arousing emotions, like fear and anxiety [14, 36]. These emotions have been found to be more powerfully elicited in more immersive VEs [13, 14], whereas less arousing emotions like relaxation and happiness are not affected by the immersiveness of a VE. Two studies have reported that presence impacts sadness elicitation [14, 37, 56], and Gilpin et al. [2021] have determined that the activating subtype of sadness is more likely induced when presence is high. This makes Freeman's [2005] theory of emotion and presence important in the attempt to elicit sadness in VR. They theorise that emotion and presence are only related for arousing stimuli, and for stimuli to be arousing, it must be perceived as personally relevant and significant [13, 14].

It is broadly believed that presence can be maximised in two ways. Firstly, a system with fidelity that is high enough to be indistinguishable from reality (i.e., perfect immersion) would probably induce the most presence. While intuitive, this idea may be difficult to achieve. The second way is that presence is a result of how immersion and the human perceptual system interact [25]. This requires one to investigate the human perceptual system to understand what makes humans convinced of reality. If evaluated correctly, one could match a system's displays and interactivity to those required by human perceptual and motor systems.

An element that may contribute to presence, and thus how effective a VE is, both in terms of its effect on a sense of presence and emotion elicitation [16, 60], is how users can control and interact with a VE. In a study comparing voice control in interactive VR to traditional "point and click" interfaces, Osking and Doucette [2019] argued that older selection systems create a jarring interruption in one's sense of presence, especially when they interrupt a narrative that one is trying to follow. Instead, a voice control dialogue system allows participants to remain grounded in their illusion of presence. The results of their study support this hypothesis. Beyond this study, which investigated interactivity in relation to presence, the effect of interactivity has not been studied much. This could be because movement in a VE is strongly related to cybersickness symptoms [5], so many developers opt to remove the need for users to move around. The effect that haptic feedback has on levels of presence has also been considered, as it may lead to greater effect from a VE [38].

2.2.2. Other Design Considerations

Beyond presence, user experience must also be considered when designing an effective VE. This can be evaluated with Sutcliffe et al.'s [2004] Heuristic Evaluation for assessing VE user interfaces, which is based on work by Sutcliffe and Kaur [2000] and Nielsen's Usability Heuristics [1994]. Hardware implications and human depiction influence user experience to a great extent, and are discussed in section 3.3.

2.3 Limitations of current literature

The literature on eliciting emotions in VR has several areas that could be improved. Importantly, larger sample sizes and more thorough recordings of treatment effects are needed. This was highlighted by Lindner's [2017] review of the use of VR in exposure therapy [5, 40, 41]. Susindar et al. [2019] also note that it is helpful to include an objective measure of emotion and presence, because of the shortcomings of self-reported measures.

Besides exceptions like Gilpin et al.'s study [2021], the literature mostly consists of studies focused on multiple emotions, which may impact the measurement of emotions in isolation. For example, Rivu et al.'s [2021] findings about eliciting sadness may be distorted because their participants experienced four different emotions, which may have been conflated with one another [4]. Studying emotions in isolation is important for sadness, as it is an interacting emotion. This means that it often presents with other emotions which significantly influence its expression [39].

In the pursuit of eliciting pure sadness, obscuring the aim of emotion elicitation from participants is also important. Results may be warped when this aim is made clear [42] because of the Hawthorne effect, in which user behaviour changes when desired testing outcomes are known [43]. Ethical standards must be carefully considered when doing so, and the literature provides some examples of this which are discussed in section 4.2.

To strengthen the analysis of elicited emotions, it has also been suggested that physiological measurements be expanded. Measures of parasympathetic activation or withdrawal, which relate to heart rate, are easiest to identify emotions with, as they are most common in current literature and can be validated [10]. However, with the suggestion of subtypes of sadness, a measure of sympathetic activation should be included, as it would strengthen the understanding of the kind of sadness induced [55].

3 System Development and Implementation

This system was designed with Software Engineering best practice in mind. An agile approach was followed, as described in sections 3.1-3.4. The final system includes four scenes: a tutorial, pet store, park, and vet.

3.1 Overview of system features

The final VE used for the user study contains two levels: one allows for almost no interaction with, or response from the dog, and one allows for interaction and response. Both interaction levels follow the same deterministic narrative, as shown in the flow diagram in Appendix 1.

3.1.1 Level 1 Features

Tutorial: The user is taught how to look and move around in the VE, use the controller's D-pad to snap their viewing camera from right to left, pick up objects and throw them, and teleport to a chosen destination. To personalise the experience and increase presence, the user is prompted to modify their hand style according to sex and skin tone before exiting the tutorial.

Pet Store: The user is introduced to the dog, Bella. Bella's animation state controller can be seen under Dog 1 in Appendix 2, and shows how she follows a set of animations on the user's arrival. On a time-based signal, her cage door automatically opens, and she trots towards the user. Bella and the user are teleported to the park three seconds after she reaches the user.

Park: The user is able to watch and follow Bella as she enacts a set of determined movements, using the states depicted in

Appendix 3. The user has no interaction with Bella other than some haptic feedback when they touch her, to which she does not respond. The user is teleported to the vet scene three seconds after hearing Bella's death, to prevent them from attempting to find her, as this may trigger a trauma response.

Vet: The user hears the vet's footsteps, prompting them to look towards the wall from behind which she appears. The vet approaches the user and stops in front of them to explain that Bella has passed away. After this, the simulation fades to black.

3.1.2 Level 2 Features

Level 2 is fundamentally the same as level 1, but a few changes are introduced.

Tutorial: Lessons on stroking and calling the dog are included.

Pet Store: The user now has a choice of three dogs, each with a distinct animation controller, outlined in Appendix 2, to give them a sense of differing personalities. The user also opens the cage door of the dog they select, unlike the timed opening in level 1.

Park: Here, the user is able to interact with their dog via stroking, throwing a tennis ball, and calling it. The dog reacts to stroking by closing its eyes and whining happily. When called, it runs towards the user and looks up at them on arrival. When the tennis ball is thrown, the dog runs after it and returns to drop it at the user's feet. These interactions are blocked in certain states, like the teddy-bear collection sequence, to allow for progression of the scene and create a sense that the dog, who is still a puppy, has a mind of its own.

Vet: the vet offers the user their dog's tennis ball, which the user is prompted to take in order to end the simulation.

3.2 Requirements gathering and analysis

To identify the requirements for the VE and user studies, several sources were consulted. First, ongoing meetings with the project supervisors were used to ensure that the scope of the project remained feasible and would allow for its results to be valid. Second, feedback on the previous iteration of this project (referred to as Iteration 0) was consulted to see what could be built on and improved. Here, it is important to note that most features of Iteration 0 were changed, as beyond the micro-issues of the project, the overall VE was not deemed effective. This was largely because of the atmosphere created in Iteration 0, which was not conducive to bonding, and the choice of the dog asset, which was not like-able enough to facilitate bonding. The scope generated during the initial planning phase of this project can be summarised in three parts:

1. Redesigning and developing the VE
2. Testing the VE in a series of pilot tests
3. Running a user study to test the difference in emotion elicitation between the levels of interaction

From the scope, it was determined that a number of resources would be needed. All VR equipment and assets from the Unity asset store, purchased for development, were provided by the UCT Computer Science Department. Similarly, a VU-AMS device to measure physiological responses was provided. The developer was expected to learn how to code in Unity and C#, and to source various self-report measures, found in the literature. Finally, Microsoft Teams was used as a tool to centralise all communication and data used for the project.

3.3 Design approach

This project was the third iteration of the VE's design. Within this iteration, an agile approach was followed to ensure that feedback could be collected and implemented at every stage of the project. The resulting iterations of this project are as follows:

Iteration 1: Planning and fixing issues from Iteration 0

The Heuristic Evaluation performed in Iteration 0 was consulted to determine the most important issues to fix. This was accompanied by discussions with the project supervisors to determine other priorities. The main goal for development was to improve the atmosphere and presence of Iteration 0. To do so, Stanney and Sadowski's [2002] work on the seven dimensions that affect presence was consulted. This led to a careful consideration of the VE's duration, realism, relationships with other VEs, social interactions, ease of social interactions, and user perception of control. Because user perception of control is directly linked to the independent variable of this study, interaction, all other elements were considered control elements, and were thus designed to be as good as possible. The seventh dimension involves characteristics of the VR system itself, which are beyond the scope of this project [32, 60]. The deliverable from this iteration was a set of prioritised project goals, which informed the remaining iterations.

Iteration 2: Altering the stylistic elements of Iteration 0

In accordance with the results from Iteration 1, before altering the dog and its behaviour, the stylistic elements of the previous VE were updated to create a VE that is more conducive to user bonding and a sense of presence. Within this iteration, the dimensions of realism, relationships with other VEs and social interactions were considered. To increase realism, the proportions of most of the elements in the pet store and vet scenes were altered, and the park environment was redesigned to make the experience more pleasant and realistic for users. Iteration 0's relationship with other VEs was considered adequate, as the controls used throughout the VE are industry standard. To improve social interactions, the Iteration 0 dog asset was replaced with a new one, so that likeability and ease of interaction could be improved. The updated VE can be seen in Appendix 4.

Iteration 3: Implementing the behaviour of the dog

This iteration focused on improving social interaction. Creating a likeable and believable dog is perhaps the most important part of this system's development. VR does have some hardware limitations to consider here. There is a limit on the number of objects a VE can contain, its level of interactivity, and its resolution. Some research suggests that, when added strategically, this visual complexity benefits emotion elicitation. For example, appropriate motion in a crowd increases anxiety in public speaking simulations [5, 21], and the movement of a spider has been determined its most frightful characteristic [5, 38]. This suggests that the graphics of critical affecting elements, here the dog, need to be prioritised in terms of realism. As such, this iteration involved designing several animations to be applied to the dog in the pet store and park scene, and an animation state controller was made to ensure that the sequence of animations was believable. The three different choices of dog, as well as their animation state controllers, were also implemented.

Towards the end of this iteration, the project supervisor was invited to do a run-through of the VE. Several issues were flagged, including the speed of the dog's movement not matching its animations, several bugs in interactivity, and an incorrect offset from the ground in the park, which made the dog look like it was floating. The dog's movements in the park were also deemed unfavourable for bonding, as the dog walked away from the user too much. These issues were prioritised, and those deemed vital were corrected to create the VE used in the pilot tests.

Iteration 4: Pilot tests and finalisation of the environment

A set of pilot tests, described in section 3.5, were used to determine usability issues before the VE was finalised. This served as an informal Heuristic Evaluation.

The feedback from these tests resulted in several changes to the VE. First, design problems were considered, beginning with general user confusion. Users did not understand the tutorial instructions easily, so the instructions were updated to be clearer. It was also noted that the researcher running the simulation during tests may have to verbally guide participants through hand selection. Users were also unsure of what they were supposed to do in the pet store, so it was decided that a brief outline of the VEs narrative (excluding its ending) would be given to each participant before entering the VE. A sign was also included in level 1's pet store to indicate that the user is being introduced to Bella. In the park scene, some users got lost in the flower grove, and were not always sure what the sound effect of the car crash meant. To fix this, a teleport barrier was implemented for the grove, and the sound effect and timing of the dog's death was improved. In the vet scene, users did not know what they were waiting for. To remedy this, the vet appears sooner in the scene, and the sound of her footsteps were added to guide the user's attention to her.

Next, several disruptions of presence were identified. In the pet store, the dogs' cage doors did not open correctly, causing them to vibrate and open into the ground. The wind in the park scene also made the grass look unrealistic, as it moved too far to the left and right. Both of these issues were fixed in time for the user study. Under social interaction, the dog's animations and sounds were not correctly synced, and the vet appeared to induce an "uncanny valley effect" [5, 44]. This is because users usually respond to human characters with familiarity, especially if the human simulation is highly realistic. However, when a human simulation is human-enough, but not perfectly human, the uncanny valley effect is invoked. At this point, measured emotion changes severely, and could even be replaced with revulsion. The previous iteration of this project used a hyper-realistic avatar with complex shading for the vet, but because she did not move very naturally or speak in sync with her mouth movements, many users reported that she was 'creepy" or got a fright when seeing her. Because the VE is not at a photorealistic stage, the vet was altered to be more semi-realistic [5] by adding the sound of footsteps to make her walking movement more believable, and changing the pace of her movement to be more natural. Her shading and finer details were also toned down. Several technical flaws, including a faulty snap-left application in the tutorial, a black screen at the beginning of every park scene, a disappearing stick object, and the user's ability to teleport to the scene of the dog's death, were fixed. It was also determined that the headset should be restarted, and its cable unkinked between users to ensure minimal lagging.

Finally, some opportunities were identified to improve the prospect of users bonding with the dog. The amount of time spent with the dog in the pet store before teleporting to the park was increased, and the dog was made to play much closer to the user in the park scene. Animations were also improved to give the dog more of a personality, and it was determined that level 1 participants needed to be briefed that the dog would not respond to their actions. A detailed code-analysis of the pilot test users' responses to the VE can be found in Appendix 5.

3.4 Quality checks during development

Throughout development, several elements were prioritised to ensure a quality system. First, because the inherited project was complex and took time to understand, the code was developed with better maintainability in mind. The software is modular, as different assets and animations can be swapped into existing code while maintaining correct behaviour. The directory system is organised for high cohesion, so that it is easy to understand which tasks are related and where there are dependencies.

The code has been made more efficient by removing costly assets (such as sound files and textures) that were unnecessary in the updated VE, and by repeating texture use etc. so that fewer textures were stored and loaded, thus lowering memory requirements. The code is also portable: all non-fundamental assets can be downloaded when needed so that only fundamental assets need to be transferred between machines, decreasing transfer time. The technical fidelity of the VE was constantly interrogated to ensure that its frame rate could be kept at a constant 60 frames per second, as this matches the capture rate of the human eye and prevents lagging [13]. Lower frame rates likely reduce presence, alter emotional response, and increase the likelihood of simulator sickness [45]. It has been determined that high graphical quality in a VE is not necessary for emotion elicitation [5]. However, because of the attempt at realism made by the original code, care was taken to ensure that real-world objects were represented as believably as possible.

The code was designed to be acceptable and compatible with as many participants as possible. Using the Steam VR engine allows the VE to be personalised according to a participant's height and position, and participants can customise their skin tones and hand styles. The language used throughout the VE was also deemed accessible through the pilot tests.

To ensure dependability and security, care was taken to ensure that the hardware could perform at full capacity. The headset cable was unkinked after every use, and the controllers were always fully charged. All users were briefed about the equipment (eg. They were shown the mesh indicating the wall, and were warned about smacking the controllers against it), and the room sensors were monitored to ensure that the room boundaries were correct. It is unlikely that a malicious attack would be made on this software as it is not public, however, important modifiers were kept private to ensure that the system remained true to its purpose regardless of external modification during its run.

3.5 Testing

Throughout the project, various testing methods were employed. Initially, corrective regression testing identified errors in Iteration 0's functionality. These errors were documented to be corrected or discarded throughout the development life cycle.

As development was carried out, unit tests were performed to ensure that the code remained functional. An incremental approach to integration testing was used to ensure that the updated software modules were integrated correctly with Iteration 0. As changes to the code became more complex, integration tests failed often, because Iteration 0 was very inter-dependent, so it was hard to implement a change without affecting the program elsewhere. This led to repeated examinations of, and adjustments to, the entire system, which led to a more loosely coupled system.

To test for errors or instability that may occur from different user input, repeated run-throughs of the VE with different user interactions were performed. To externally validate the system, the project supervisor performed an informal alpha test to pick up on obvious system flaws. A series of beta pilot tests were then conducted to ensure that, beyond its affectiveness, the VE was fully useable. Six members of the Computer Science Honours class performed this usability testing. They were informed of the nature of the VE and subjected to prescreening tests, then asked to identify features that did not work, areas of confusion, and how the dog could be more likeable. They completed the post-immersion questionnaires outlined in 4.2 for timing purposes. Their feedback determined the changes made in Iteration 3.

In future iterations of this project, or projects like it, a more extensive pilot test stage is advisable, so that usability issues can be eliminated before the actual experiment is carried out. It is also recommended that the alpha test is performed a suitable length of time after the pilot tests and the changes as a result, and takes the form of a rigorous Heuristic Evaluation. This should result in a more thorough and effective completion of the VE. Time constraints prevented this in this study.

4 Experiment Design and Execution

4.1 Participants

A user study of 43 participants was completed for this experiment. The participant population was made up of undergraduate UCT students who responded to the study advertisement and met its exclusion criteria. The exclusion criteria were chosen, according to the literature, to ensure that participants who are predisposed to severely negative reactions would be excluded from the study. To participate, users had to score below 45 on the State-Trait Anxiety Inventory (STAI), below 3 on the Primary Care PTSD Screen for DSM-5 (PC-PTSD-5), and below 21 on the Beck Depression Inventory–II (BDI-II). Appendix 6 contains a description of these and why they were chosen. Users also had to indicate that they were not afraid of dogs, using a 10-point scale to rate their fear. If they scored above 5 on this scale, they were excluded.

The resulting participant pool was diverse, with few dominant traits to bias the data. The only notable trait of this pool was that the majority were from the UCT Psychology Department. The pool was randomly allocated to either level 1 or 2. No students with relationships to the researchers were included to ensure no conflicts of interest. Below is a table to show that there was no significant difference in participants between level 1 and 2:

Table 1: between condition differences (n=43)

Variable	Group		
	Level 1 (n=21)	Level 2 (n=22)	P-value (one-tailed t test)
Year of study M (SD)	2.0476 (0.212426)	2.2273 (0.254224)	0.29617
Faculty			
Commerce	2	1	
Engineering	1	3	
Health Sciences	1	0	
Humanities	15	15	
Law	1	0	
Science	1	3	
VR experience M (SD)	1.333333 (0.210819)	1.318182 (0.258085)	0.29617

[1]

4.2 Ethical considerations

Because this study involved human participants, ethics clearance needed to be obtained from the UCT Psychology Department. The factors considered to ensure ethical practice are outlined below.

4.2.1 Justification for Masking the Study's Purpose

Studies relating to emotion elicitation are particularly vulnerable to the Hawthorne effect [43]. Disclosing such studies' true purposes has been shown to skew results [42]. To avoid this, the study was masked by telling users that it aimed to develop a platform in which people can learn to train dogs.

To ensure that this masking adhered to research ethics standards, the American Psychological Association's *Criteria for Valid Deception of Research Participants* [2002] was consulted. In accordance, we first ensured that the deception involved was justified by its significant contribution of information to the field of psychology. Given that there is no alternative way to complete this study with adequate validity, this benefit was considered justification enough.

Next, the exclusion criteria were used to ensure that the research would not cause severe distress to participants, and it was made clear that participants could stop the simulation at any point should simulator sickness, or other distress, occur. Finally, all participants had a post-experiment debriefing in which the true purpose of the study was verbally explained and written in a document that they could take home. They were invited to ask questions to ensure that true consent to use the participant's data was obtained. No participants experienced simulator sickness, and

[1] VR experience was estimated by giving each participant a rating from 0-2 on their VR and gaming experience (0=none, 1= non-interactive/limited experience, 2=interactive experience)

participants who indicated distress were given a space to discuss the experiment with the researcher, and read through the debriefing form with them, until the researcher was satisfied that they were either calm or could access help if needed.

4.2.2 Other considerations

We prioritised having minimal risks because the study was unlikely to directly benefit participants, beyond a small participation payment. This payment was not large enough to be considered coercive and followed standard UCT research practice. Free and informed consent was collected at the beginning of each test with a consent form signed by the participant and researcher. Care has been taken to ensure that this report interprets results according to the data collected without alterations or false reports. All study-related data was anonymised and will not be used beyond this study.

4.3. Measurement

Sadness presents itself physiologically and behaviourally, and can be recorded through physiological identifiers and via qualitative self-report [3, 50]. Most studies attempt to draw conclusions about emotion elicitation in VR using one of the two methods, rather than combining them. For a thorough and triangulated approach, this study uses objective physiological data and subjective qualitative measures. This means that the results are a careful consideration of a series of metrics, all deemed suitable by previous work. All measurements were taken in the same order, and subjective measures were taken directly after immersion. The researcher was available to answer questions relating to the questionnaires, but did not interfere (i.e., did not read over the participants shoulder, or encourage them to answer in a specific way). This minimised bias from the researcher. Below is an explanation of the measurements recorded during the procedure:

4.3.1 Physiological

To investigate the subtype of sadness elicited in either level, and how extreme the experience of sadness was, we followed the study performed by Gilpin et al. [2021] and used the Biopac MP160 system to measure heart rate (HR), respiration rate (RR), respiratory sinus arrhythmia (RSA), and skin conductivity level (SCL) during immersion. To identify the effect the VE had on participants, three time periods were recorded in each study: the time for which the user was in the pet store (PS), the time for which the user was in the park (P), and the time for which the user was at the vet (V). PS was deemed unnecessary to analyse as it did not include any emotional triggers related to sadness. P is believed to indicate the physiological effect of the death scene on the user's emotion. V is believed to indicate how the user felt in the aftermath of their dog's death.

Two baseline measures were taken for all four physiological measures. The first, BO, represents an out of immersion baseline. It is the average reading of the participant while in the neutral experiment room, taken over two minutes. BO was taken at least three minutes after the electrodes were first attached to the participant to allow for calming. The second, BI, represents an in-immersion baseline. It is the average reading of the participant for the duration of their tutorial. Two measures were taken in case there was a difference between baseline measurements in and out of immersion. However, there was no statistical difference found between the two measures (see Appendix 7). This meant we chose to use BI throughout our presentation of results.

Previous research meant we expected to see an activating pattern of sadness, characterised by an increase in HR (which indicates parasympathetic withdrawal), RR, and SCL (which indicates arousal), and a decrease in RSA (which indicates parasympathetic activity) [55]. Unfortunately, because of technical problems with the VU-AMS device, some data was corrupted and could not be used for analysis. This largely affected level 1, so statistical analysis has been adjusted accordingly.

4.3.2 Qualitative

Qualitative measures were taken post-immersion. The participants were asked to record their primary emotion in the final moments of the VE with the Visual Analogue Scale (VAS). The VAS asks participants to choose the emotion they feel most strongly out of a list of emotions, and rate it from one to ten [27, 51]. This scale has been used by Riva [2007] and Freeman [2005]. It was used in place of the Primary emotion Rating (PER) used by Gilpin et al. [2021], which also asks participants to choose the primary emotion they feel after experiencing the sadness inducing stimulus, but does not ask them to rate it [52].

Similarly, the nine-point Self-assessment mannequin scale (SAM) was used to assess participants' emotion in the final scene on a dimensional visual scale. SAM is a valence-arousal model, which uses visual representations to categorise emotions as linear functions of valence and arousal [4, 45, 52]. Its third axis, dominance, indicates the level of control the participant feels they have over their emotion [4, 37]. It has been used in a study to validate that VR can be used to research emotions and emotion elicitation, given how similar its effects are to reality [4], and in a study measuring emotional response to advertising [53]. It was initially proposed as an alternative to inefficient verbal self-report measures, which is useful in our time-constrained study [53]. It is also useful in the context of the culturally diverse population pool we studied, as it is purely visual so does not rely on the understanding or cognitive comprehension of words, which often interrupts one's ability to sense how they are feeling. Rivu et al. [2021] suggest that SAM be used with objective physiological measurements to record substantial results, as done in this study.

Participants were asked to fill out the Discrete Emotions Questionnaire (DEQ), a robust discrete measure of possible co-present emotions, with the entire simulation in mind. The DEQ is fairly novel, and was designed to detect eight distinct emotions: anger, anxiety, desire, disgust, fear, happiness, relaxation, and sadness. It was designed in response to the tendency of popular self-report measures to result in null effects, discounting some emotions that may be present [54]. It has been validated with four studies that directly compared it to the widely used Positive and Negative Affect Schedule (PANAS), and was found to be more sensitive at detecting self-reported emotions [54]. This is helpful, because emotions are only probabilistically linked to their measurement, so it seems inappropriate to assert that a null effect from a questionnaire like PANAS means that emotion was not involved in a situation of interest. It is also preferable to PANAS in theory because it is simple, quick to fill out, and was created from words used by everyday people when describing their emotions. This should make it understandable to English speakers, but we observed some confusion about certain words, such as revulsion. This is probably because the questionnaire was created in America, where words considered common may not be common to the average South African.

The Brief Presence Questionnaire (BPQ) was also used to gauge user's sense of presence. Bouchard [2010], Bouchard et al. [2008], and Diemer et al. [2015] have all validated the BPQ, which involves simply asking participants "On a scale of 0 to 100, how much did you have the impression of really being there in the virtual environment?" [14, 16]. An adapted version of this, where participants were asked to use a scale of one to ten, was applied in this study. This was used instead of other popular questionnaires like the UCL Presence Questionnaire [23] and the Independent Television Company Sense of Presence Inventory [13, 55] because they were too long for the time allocated to each participant. While there is no evidence to suggest that this makes the presence values less valid, we do consider that our measure of presence may not be thoroughly accurate. This is acceptable within the context of this experiment, as immersion was a controlled variable, meaning that presence would not have been influenced by anything other than subjective response.

A shortcoming in the design of this experiment is that no baseline subjective measures were taken. Moreover, no subjective measures were taken during immersion, which is not considered best practice [22, 23, 25]. However, when combined with the physiological measurements and each other, stronger assumptions can be made. The option to use less measures and take them more than once, or increase the length of the test period was considered. However, the possible loss of participant sign-up because of time constraints, and the limitations of relying on one or two subjective measures, were deemed a strong enough risk to prioritise more measurements being taken in a shorter period of time.

4.4 Procedure

Because most work done on eliciting emotions in VR has not focused on sadness, this experiment design comprises a combination of best practice in previous work on eliciting emotions in VR in general. The experiment room used was cleaned and cleared before each participant to ensure that only the necessary equipment was visible, to avoid and control response to the room itself. See Appendix 8 for the experiment procedure.

4.5 Data management and statistical analysis

To analyse the data, descriptive statistics were generated for each sample. The kurtosis and skewness were evaluated to estimate normal distribution. If both lay between -2 and 2, the data was estimated to be normally distributed. If not, the data was inspected for outliers beyond three standard deviations above or below the mean. If such outliers were found, they were discarded, and the samples were retested for normality. If not, the data was regarded as not normally distributed.

A t-test was used to compare level 1 samples to level 2 samples if both were normally distributed. If sample sizes and standard deviation obviously differed, this test was performed assuming unequal variances [49]. If the sample sizes and variances were reasonably similar, the t-tests were performed assuming equal variance. If both samples were not normally distributed, a Mann-Whitney U test was used, as this type of non-parametric test compares the differences between two small samples when they are not normally distributed [48]. In the instance where one sample was normally distributed and the other not, both a t-test and Mann-Whitney U test were used and interpreted.

5 Results and Discussions

Each measurement's statistical analysis and other noteworthy findings are outlined below. It can be assumed that both samples (from level 1 and 2) are normally distributed unless stated otherwise, which makes analysis of the samples' means and modes more valid. According to the results of statistical analysis performed, it can also be assumed that no statistically significant difference was found between the level 1 and level 2 samples at $p < 0.05$, unless otherwise stated. Summary statistics and test summaries can be found in Appendix 9.

5.1 Summary of physiological results

Table 2: summary of tests for significant difference

Test description (variance)	L1 sample size	L2 sample size	Test type	P (T<=t) one-tail /P-value	P (T<=t) two-tail /P-value	Significance?
HR P-BI (unequal)	13	21	t-Test	0.3226	0.6452	NO
HR V-BI (unequal)	13	21	t-Test	0.0481	0.0962	YES
RR P-BI (unequal)	14	20	t-Test	0.1708	0.3416	NO
RR V-BI (unequal)	14	20	t-Test	0.1999	0.3998	NO
RSA P-BI	14	19	Mann-Whitney U	0.2981	0.5961	NO
RSA V-BI (unequal)	14	19	t-Test	0.0540	0.1080	NO
SCL P-BI (unequal)	6	20	t-Test	0.0030	0.0060	YES
SCL V-BI (unequal)	6	20	t-Test	0.0095	0.0191	YES

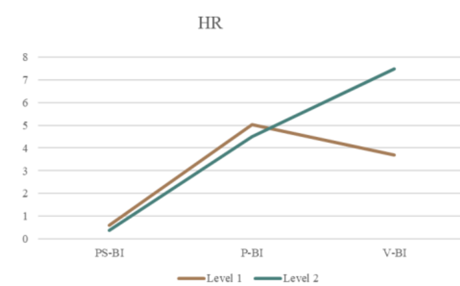


Figure 1: mean change in HR across the different levels

The mean change found in HR shows that both levels experienced an overall increase in HR, which is associated with an activating pattern of emotion. This could be because of general arousal from the VE, and specifically because of an anxiously aroused response to the dog's death scene (P). However, level 1's HR decreased after the dog's death (V), which is indicative of deactivating sadness, while level 2's continued to increase. The differences in HR during V were statistically significant according to a one-tail t-test. This indicates that, while the VE was activating, the sadness induced by the dog's death was deactivating for level 1, and activating for level 2.

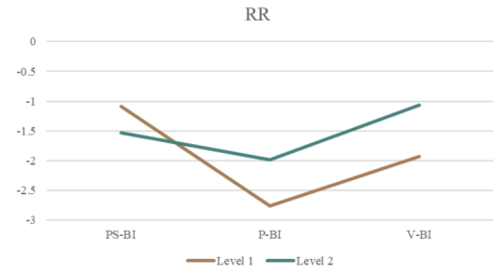


Figure 2: mean change in RR across the different levels

In contrast, the mean change in RR suggests that both levels experienced an overall decrease in RR, which is associated with deactivating sadness. However, both level's RR increased after the dog's death, indicating that the sadness induced by the death may have been activating in both cases.

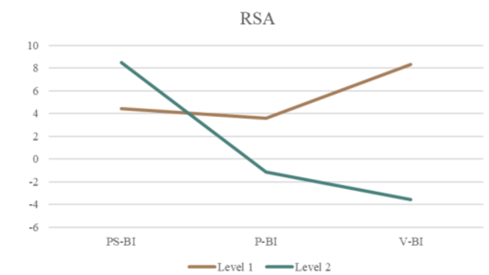


Figure 3: mean change in RSA across the different levels

Here, the initial decrease in RSA is related to activating sadness. However, level 1 experienced an increase in RSA after the dog's death, which is related to deactivating sadness. Level 2 had a clear decreasing trajectory, suggesting that their state became more activated as a result of the dog's death. Both level's P-BI samples were not normally distributed, with a kurtosis of 2.1599 in level 1 and 2.3221 in level 2. These kurtosis values are fairly close to normal, and are also fairly close to one another, however, they suggest that analysis of the mean values here should not be deemed conclusive.

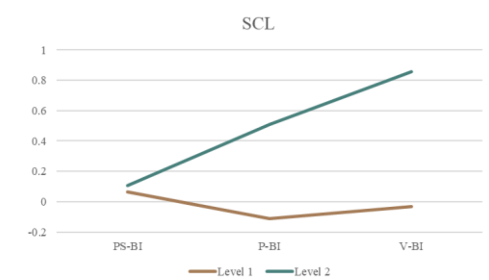


Figure 4: mean change in SCL across the different levels

Here, a clear distinction between the levels' experiences in the VE is suggested. Level 1 experienced an overall decrease in SCL, which is again related to deactivating emotion. However, it did increase a small amount post death. Level 2 had a clear increasing trajectory, suggesting once again that their state became more activated as a result of the dog's death. The difference in SCL was found to be significant according to one and two tailed tests in both the park and vet environments.

5.2 Summary of self-report results

Table 3: summary of tests for significant difference

Test description (variance)	L1 sample size	L2 sample size	Test type	P (T<=t) one-tail /P-value	P (T<=t) one-tail /P-value	Significance?
VAS sadness (equal)	13	12	t-Test	0.1796	0.3592	NO
DEQ sadness (equal)	21	22	t-Test	0.4650	0.9300	NO
SAM Row 1	21	22	Mann-Whitney U	0.4801	0.9601	NO
SAM Row 2 (equal)	21	22	t-Test	0.2102	0.4204	NO
SAM Row 3 (equal)	21	22	t-Test	0.4341	0.8683	NO

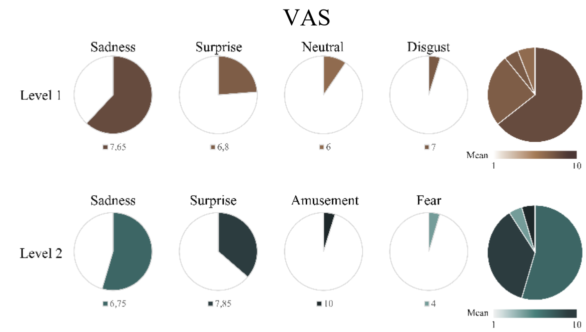


Figure 5: the number of users who chose the indicated emotion, and their mean rating of the emotion

Unsurprisingly according to the physiological changes noted above, both levels had sadness reported as the participants' primary emotion by the majority. It is interesting to note that the number of people who chose sadness, and the level of sadness felt, was slightly lower on average in level 2, and that surprise was slightly higher. It makes sense that surprise, an activating emotion, would be more keenly felt by level 2 as they, on average, were more emotionally activated by the VE. The level of surprise was not normally distributed in level 1, but when comparing level 1's mode of 4 with level 2's mode of 8, the mean difference does seem to indicate a valid trend.

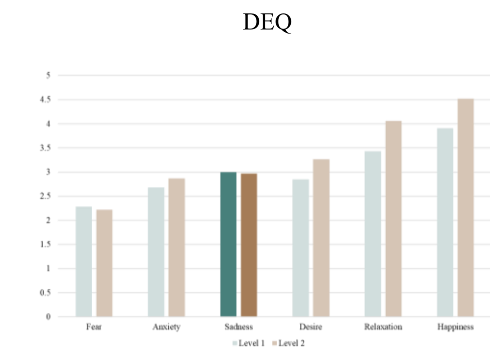


Figure 6: the difference between the mean rating of each emotion category in the DEQ

The results of the DEQ again show that participants in both levels experienced sadness as a result of the VE, and that level 2 participants rated their experience of sadness lower. Anger and disgust were left out of analysis because they were reported very rarely and with a high variance.

Interestingly, level 1 only beat level 2's ratings of fear and sadness. Perhaps elevated activation in level 2 allowed them to experience other emotions more intensely, specifically positive emotions which were reported much more highly. It could also be the case that participants simply had more fun in level 2, where they felt they had a purpose, as opposed to level 1 where they largely acted as observers.

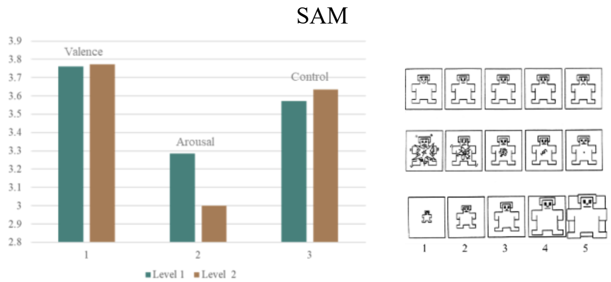


Figure 7: mean SAM ratings alongside SAM scale

The SAM results are quite anomalous. They suggest that participants in level 1 felt less negatively (had a lower valence) than those in level 2, and that level 1 participants had a higher state of arousal and a lower sense of control. Level 2's valence results were not normally distributed, which may make a mean-based assumption less accurate. However, the measures of control and arousal do come from a normally distributed sample.

According to the modes for both levels, participants most commonly reported 4 for valence and 5 for control, indicating that they felt similarly unhappy and in control. However, the mode for arousal was 4 for level 1, and 2 for level 2. This indicates that level 1 participants felt more excited, or stimulated, by their emotions. This strongly contrasts the physiological data.

Other than statistical outliers, this could be because of a lack of understanding of the visual scale – many participants asked what it meant – but it may also show the limitations in the subjective measure, because it was taken after immersion so may indicate how participants were feeling then, rather than during immersion. Of course, it is possible that the other data is misrepresenting the sadness elicited, but this is unlikely given the data's consistency.

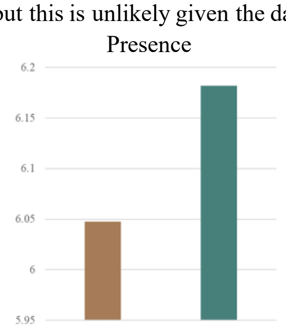


Figure 8: presence in relation to the sadness type elicited

Higher presence in level 2 is expected according to the literature: high arousal has been correlated with an increase in presence. The sample's modes support this result, with level 1's mode being a presence rating of 3, and level 2's being 8. Quantitative data collected during immersion also suggests that presence was higher in level 2, because a good indication of presence is when people respond to a VE in the same way as they would to real-life [51], and more users in level 2 spoke to the dog in a typical "puppy" voice than in level 1.

5.3 Discussion of results

The first result to address is the lack of statistical significance in some comparisons between level 1 and level 2, which suggests that there may be very little difference between the two levels. This could in large part be because of the small sample size. However, assuming that interaction does affect sadness elicitation, the results from level 2 may have been dampened for two other reasons. Firstly, there was a steep learning curve in terms of how to interact with the dog, so the interaction may not have been engaged with properly. Secondly, the VE is not developed for maximum presence and still has room for improvement, which may have been more evident in level 2.

However, despite the lack of statistical significance, some clear trends arise from the data that are consistent with the literature [10, 55]. All results point towards sadness being elicited in both levels, with the intensity being slightly higher in level 1 at a descriptive level. However, this does not necessarily indicate that level 1 induced sadness more effectively. Rather, and especially in light of the physiological results, it seems that the different levels of interaction induced different subtypes of sadness.

While overall it seems that Gilpin [2021] and Kreibig's [2007] analysis on VR generating an activating pattern of sadness is correct, the results of this research indicate that low levels of interaction in a VE elicit a deactivating pattern of sadness, while high levels induce an activating pattern. Given the results, it appears that increased interaction creates a sense of agency and accountability. In response to the dog's death, level 2's physiological measures largely pointed towards activating sadness, which usually occurs as a result of failure. This suggests that an increase in interaction in VR may cause users to feel that they are more responsible for the events in the VE, or that they are playing a game with the goal of looking after, and bonding with, the dog. Thus, when the dog dies, they feel they have failed, or lost the game. Users who reported some of the highest levels of sadness in level 2 expressed guilt in the debriefing conversation because they did not save the dog. Conversely, users who could not interact with the dog simply experienced its death as a loss, because they felt much less agency, so did not feel as though they had failed. Based on the outcomes of this research, it is suggested that interaction in VR does not necessarily affect the intensity of sadness created, but rather the type.

5.4 Discussion of anomalies

There were several anomalous results that must be discussed. Firstly, two participants with some of the highest combined levels of self-reported sadness in level 1 revealed that they had lost their dogs in the same way. Unfortunately, both participants had erroneous physiological readings, so this cannot be added to analysis. However, they both had high presence ratings, of 8 and 10. This strengthens the literature's assumption that stories of personal significance result in greater emotional effect [13, 14].

Few users reported that they did not bond with the dog (5 in level 1 and 4 in level 2). Most of these outliers reported that they did not feel that they had enough time to bond with it, or that it did not feel real. In level 1, some participants expressed that they would have bonded better if there had been more interaction. The character development of the dogs may also have been slightly greater in level 2, as each dog appeared to have a distinct personality. Interestingly, thirteen level 2 participants selected the third dog, Luna, and only two selected the dog used in level 1, Bella. The effect of this should be considered moving forward.
Despite the overall trend of sadness, some participants indicated emotions other than sadness in their VAS responses. It is hypothesised that surprise was the second most reported emotion because the study was masked, and that because level 2 participants were highly stimulated, they experienced surprise to a greater extent. Two level 1 users reported that they felt neutral at the end of the simulation, which makes sense in the context of a lower level of arousal, especially considering that both users reported that they had not really bonded with the dog. One level 1 user reported disgust, and seemed put off by the masked nature of the study. One level 2 user reported a low level of fear, which they described as "a level of concern" in the debriefing discussion. This could indicate that they were invested in the dog's well-being. Another reported amusement as they had found the VR experience very entertaining. However, they did indicate that they were saddened in other self-report measures.
A final interesting result is outliers in the presence measurement of level 2. Several users found the increased interaction detrimental to their sense of presence. This appears to have been because they were struggling to remember how to use the controls after the quick tutorial, which broke presence, or the interactions may have broken the illusion because they were not convincing enough (see Appendix 10). In future work, it would be interesting to test this system with more levels of interaction, to see if these outliers are specific to this VE and population, or if there is a cap on the amount of interaction used before presence is broken.

6 Conclusions and Future Work

Overall, this study has achieved its aims of contributing to our understanding of how emotion elicitation works in VR. The VE produced was largely successful in eliciting sadness, and the triangulated approach to measuring this sadness allowed for thorough analysis of its effect. In response to the research question, the results of the user study suggest that low levels of user interaction induce a deactivating pattern of sadness, while high levels induce an activating pattern. Moving forward, it is suggested that these results be validated by repeating a similar experiment with greater levels of control. This would include a VE that is more developed and geared towards maximising presence, a better training set up so that lack of experience does not influence the results, a longer simulation, more consideration of the effect of the dog's character development, and a larger participant pool.

REFERENCES

[1] Lisa Feldman Barrett. 2020. How emotions are made: The secret life of the brain, S.l.: PICADOR.

[2] David R. Pratt, Michael Zyda, and Kristen Kelleher. 1995. Virtual Reality: In the Mind of the Beholder: Computer 28, 7 (July 1995), 17–19

[3] Ivan Alsina-Jurnet, Cristina Carvallo-Beciu, and José Gutiérrez-Maldonado. 2007. Validity of virtual reality as a method of exposure in the treatment of test anxiety. Behavior Research Methods 39, 4 (2007), 844–851. DOI:http://dx.doi.org/10.3758/bf03192977

[4] Radiah Rivu, Ruoyu Jiang, Ville Mäkelä, Mariam Hassib, and Florian Alt. 2021. Emotion elicitation techniques in virtual reality. Human-Computer Interaction – INTERACT 2021 (2021), 93–114. DOI:http://dx.doi.org/10.1007/978-3-030-85623-6_8

[5] 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

[6] Ville Mäkelä et al. 2020. Virtual field studies. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (2020). DOI:http://dx.doi.org/10.1145/3313831.3376796

[7] Ewa Siedlecka and Thomas F. Denson. 2018. Experimental methods for inducing basic emotions: A qualitative review. Emotion Review 11, 1 (2018), 87–97. DOI:http://dx.doi.org/10.1177/1754073917749016

[8] Krista Lauwerijssen. 2008. Sadness. Bachelor's Thesis. Tilburg University, Tilburg, Netherlands.

[9] C. Z. Stearns. 1993. Sadness. In M. Lewis, & J.M. Haviland (Eds.), Handbook of Emotions (pp. 547-562). New York: The Guilford Press

[10] Sylvia D. Kreibig. 2010. Autonomic nervous system activity in emotion: A Review. Biological Psychology 84, 3 (2010), 394–421. DOI:http://dx.doi.org/10.1016/j.biopsycho.2010.03.010

[11] Mariko 11 and Naoto Suzuki. 2017. Is sadness only one emotion? psychological and physiological responses to sadness induced by two different situations: "loss of someone" and "failure to achieve a goal." Frontiers in Psychology 8 (2017). DOI:http://dx.doi.org/10.3389/fpsyg.2017.00288

[12] C. Barr-Zisowitz. 2000. "Sadness"- Is there such a thing? In M. Lewis, & J.M. Haviland- Jones, (Eds.), Handbook of Emotions (pp. 607-622). New York: The Guilford Press.

[13] Jonathan Freeman, Jane Lessiter, Katherine Pugh, and Ed Keogh. 2005. When presence and emotion are related, and when they are not. In Proceedings of the Conference at Presence 2005, London. Available at: http://www.temple.edu/ispr/prev_conferences/

[14] Julia Diemer, Georg W. Alpers, Henrik M. Peperkorn, Youssef Shiban, and Andreas Muhlberger. 2015. The impact of perception and presence on emotional reactions: A review of research in virtual reality. Frontiers in Psychology 6 (2015). DOI:http://dx.doi.org/10.3389/fpsyg.2015.00026

[15] Giuseppe Riva et al. 2007. Affective interactions using virtual reality: The link between presence and emotions. CyberPsychology & Behavior 10, 1 (2007), 45– 56. DOI:http://dx.doi.org/10.1089/cpb.2006.9993

[16] Stéphane Bouchard, Julie St-Jacques, Geneviève Robillard, and Patrice Renaud. 2008. Anxiety increases the feeling of presence in virtual reality. Presence: Teleoperators and Virtual Environments 17, 4 (2008), 376–391. DOI:http://dx.doi.org/10.1162/pres.17.4.376

[17] Barbra Zupan and Duncan R. Babbage. 2016. Film clips and narrative text as subjective emotion elicitation techniques. The Journal of Social Psychology 157, 2 (2016), 194–210. DOI:http://dx.doi.org/10.1080/00224545.2016.1208138

[18] Jennifer S. Lerner and Dachter Keltner. 2001. Fear, anger, and risk. Journal of personality and social psychology, 81(1), 146

[19] Norbert Schwarz and Gerald L. Clore. 1983. Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. Journal of Personality and Social Psychology 45, 3 (1983), 513–523. DOI:http://dx.doi.org/10.1037/0022-3514.45.3.513

[20] Gerardo Herrera, Rita Jordan, and Lucí Vera. 2006. Agency and presence: A common dependence on subjectivity? Presence: Teleoperators and Virtual Environments 15, 5 (2006), 539–552. DOI:http://dx.doi.org/10.1162/pres.15.5.539

[21] Sandra Poeschl and Nicola Doering. 2014. Virtual speech anxiety training — effects of simulation fidelity on User Experience. 2014 IEEE Virtual Reality (VR) (2014). DOI:http://dx.doi.org/10.1109/vr.2014.6802073

[22] Mel Slater and Sylvia Wilbur. 1997. A framework for immersive virtual environments (five): Speculations on the role of presence in virtual environments. Presence: Teleoperators and Virtual Environments 6, 6 (1997), 603–616. DOI:http://dx.doi.org/10.1162/pres.1997.6.6.603

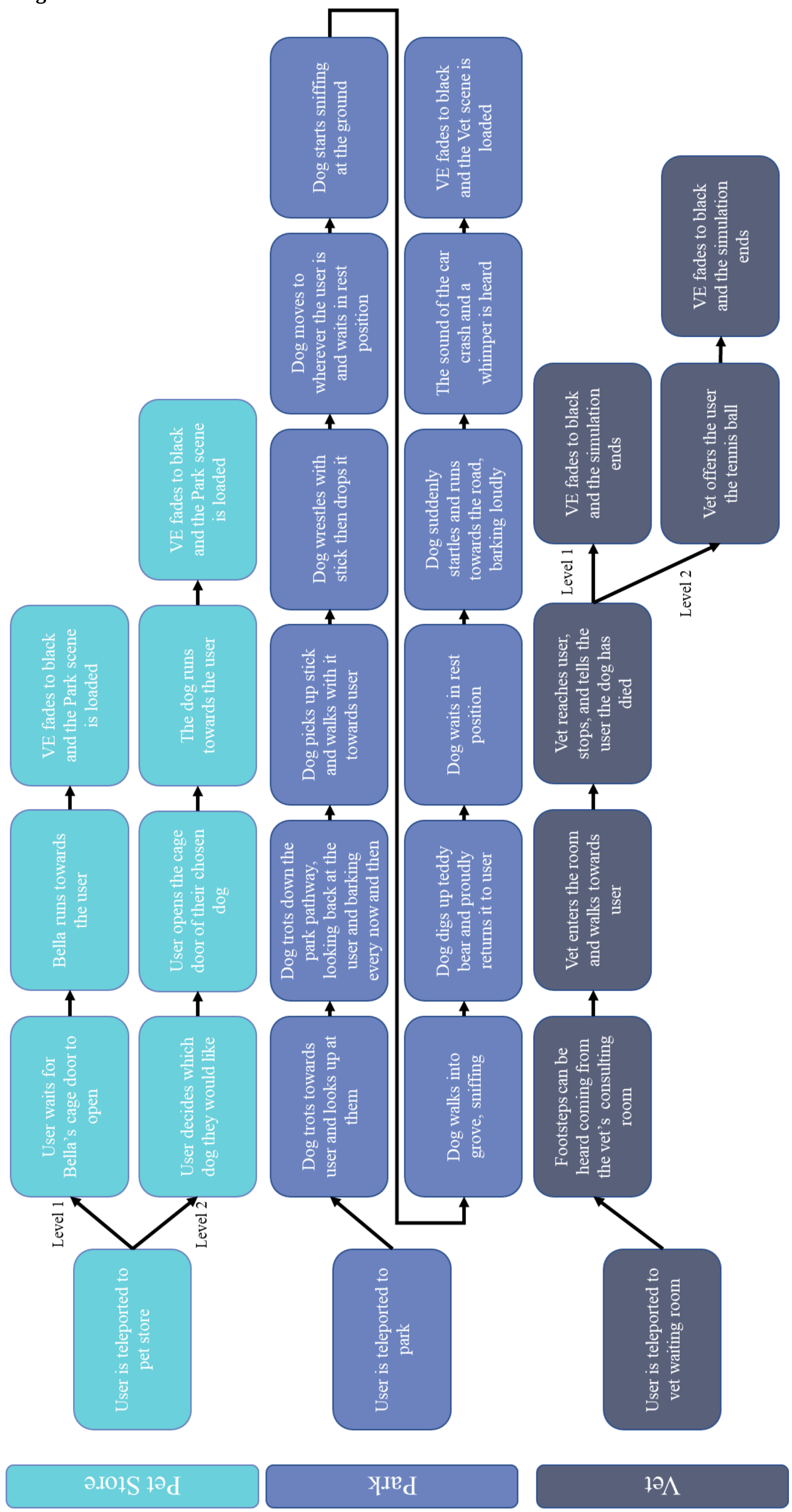
[23] Mel Slater. 1999. Measuring presence: A response to the Witmer and singer presence questionnaire. Presence: Teleoperators and Virtual Environments 8, 5 (1999), 560–565. DOI:http://dx.doi.org/10.1162/105474699566477

[24] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. 2001. The experience of presence: Factor analytic insights. Presence: Teleoperators and

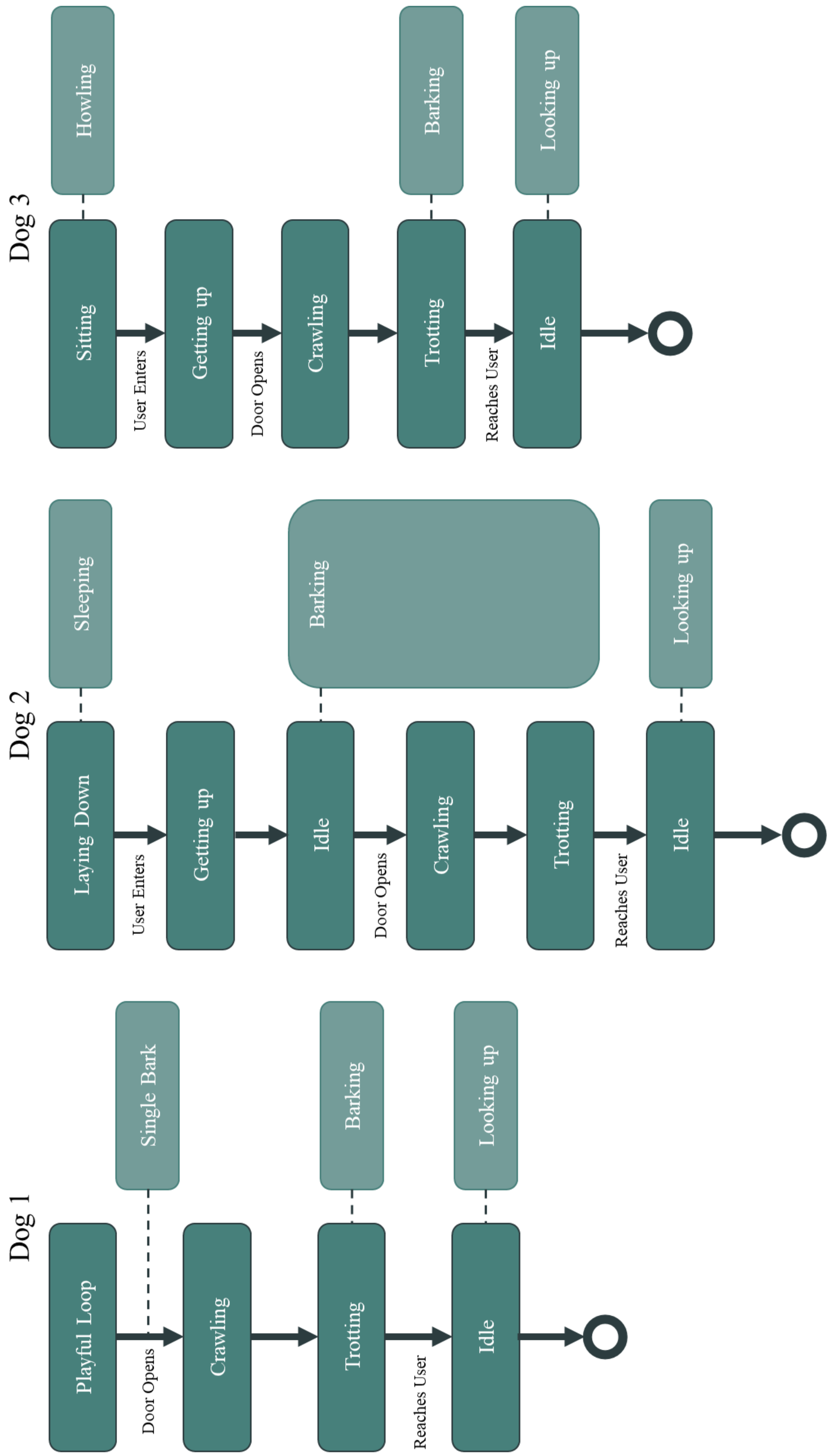
Virtual Environments 10, 3 (2001), 266–281. DOI:http://dx.doi.org/10.1162/105474601300343603
[25] Mel Slater and Anthony Steed. 2000. A virtual presence counter. Presence: Teleoperators and Virtual Environments 9, 5 (2000), 413–434. DOI:http://dx.doi.org/10.1162/105474600566925
[26] C. Botella et al. 1999. Differences on presence and reality judgment using a high impact workstation and a PC workstation. CyberPsychology & Behavior 2, 1 (1999), 49–52. DOI:http://dx.doi.org/10.1089/cpb.1999.2.49
[27] M. Carmen Juan and David Pérez. 2009. Comparison of the levels of presence and anxiety in an acrophobic environment viewed via HMD or cave. Presence: Teleoperators and Virtual Environments 18, 3 (2009), 232–248. DOI:http://dx.doi.org/10.1162/pres.18.3.232
[28] Merel Krijn, Paul M.G. Emmelkamp, Roeline Biemond, Claudius de Wilde de Ligny, Martijn J. Schuemie, and Charles A.P.G. van der Mast. 2004. Treatment of acrophobia in virtual reality: The role of immersion and presence. Behaviour Research and Therapy 42, 2 (2004), 229–239. DOI:http://dx.doi.org/10.1016/s0005-7967(03)00139-6
[29] Wijnand Ijsselsteijn, and Giuseppe Riva. 2003. Being there: The experience of presence in mediated environments. In G. Riva, F. Davide, & W. A. Ijsselsteijn (Ed.), Being There. Concepts, Effects and Measurements of User Presence in Synthetic Environments (p. 3- 17).Washington, DC: IOS Press
[30] B. K. Wiederhold, and M. D. Wiederhold, M. D. (1999). Clinical observations during virtual reality therapy for specific phobias. Cyberpsychology and Behavior, 2(2), 161-168.
[31] M. Michaud, Stéphane Bouchard, Stéphanie Dumoulin, and X. W. Zhong. 2004. Manipulating presence and its impact on anxiety. Presented at the Cybertherapy Conference 2004, San Diego, January 10-12 2004
[32] Geneviève Robillard, Stéphane Bouchard, Thomas Fournier, and Patrice Renaud. 2003. Anxiety and presence during VR immersion: A comparative study of the reactions of phobic and non-phobic participants in therapeutic virtual environments derived from computer games. CyberPsychology & Behavior 6, 5 (2003), 467–476. DOI:http://dx.doi.org/10.1089/109493103769710497
[33] Matthew Price and Page Anderson. 2007. The role of presence in virtual reality exposure therapy. Journal of Anxiety Disorders 21, 5 (2007), 742–751. DOI:http://dx.doi.org/10.1016/j.janxdis.2006.11.002
[34] Thomas D. Parsons and Albert A. Rizzo. 2008. Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: A meta-analysis. Journal of Behavior Therapy and Experimental Psychiatry 39, 3 (2008), 250–261. DOI:http://dx.doi.org/10.1016/j.jbtep.2007.07.007
[35] Yun Ling, Willem-Paul Brinkman, Harold T. Nefs, Chao Qu, and Ingrid Heynderickx. 2012. Effects of stereoscopic viewing on presence, anxiety, and Cybersickness in a virtual reality environment for public speaking. Presence: Teleoperators and Virtual Environments 21, 3 (2012), 254–267. DOI:http://dx.doi.org/10.1162/pres_a_00111
[36] Valentijn T. Visch, Ed S. Tan, and Dylan Molenaar. 2010. The emotional and cognitive effect of immersion in film viewing. Cognition & Emotion 24, 8 (2010), 1439–1445. DOI:http://dx.doi.org/10.1080/02699930903498186
[37] R.W. Picard, E. Vyzas, and J. 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:http://dx.doi.org/10.1109/34.954607
[38] Graham C.L. Davey. 1991. Characteristics of individuals with fear of spiders. Anxiety Research 4, 4 (1991), 299–314. DOI:http://dx.doi.org/10.1080/08917779208248798
[39] Alistair Sutcliffe and Brian Gault. 2004. Heuristic evaluation of virtual reality applications. Interacting with Computers 16, 4 (2004), 831–849. DOI:http://dx.doi.org/10.1016/j.intcom.2004.05.001
[40] Sarah Page and Matthew Coxon. 2016. Virtual reality exposure therapy for anxiety disorders: Small samples and no controls? Frontiers in Psychology 7 (2016). DOI:http://dx.doi.org/10.3389/fpsyg.2016.00326
[41] Rukshani Somarathna, Tomasz Bednarz, and Gelareh Mohammadi. 2021. Virtual Reality for Emotion Elicitation - A Review.
[42] 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
[43] Berkely Rice. 1982. The Hawthorne defect: Persistence of a flawed theory. Psychology Today, 16(2), 70-74
[44] Jun'ichiro Seyama and Ruth S. Nagayama. 2007. The uncanny valley: Effect of realism on the impression of artificial human faces. Presence: Teleoperators and Virtual Environments 16, 4 (2007), 337–351. DOI:http://dx.doi.org/10.1162/pres.16.4.337
[45] 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
[46] Sahinya Susindar, Mahnoosh Sadeghi, Lea Huntington, Andrew Singer, and Thomas K. Ferris. 2019. The feeling is real: Emotion elicitation in virtual reality. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 63, 1 (2019), 252–256. DOI:http://dx.doi.org/10.1177/1071181319631509
[47] James A. Russell. 1980. A circumplex model of affect. Journal of personality and social psychology 39, 6 (1980), 1161.
[48] B Rosner and D Grove. 1999. USE OF THE MANN-WHITNEY U-TEST FOR CLUSTERED DATA.STATISTICS IN MEDICINE Statist. Med 18: 1387–1400. https://doi.org/10.1002/(SICI)1097-0258(19990615)18:11
[49] Graeme D. Ruxton. 2006. The unequal variance t-test is an underused alternative to Student's test and the Mann–Whitney U test. Behavioral Ecology 17, 4: 688–690. https://doi.org/10.1093/beheco/ark016
[50] Peter J. Freed and John J. Mann. 2007. Sadness and loss: toward a neurobiopsychosocial model. American Journal of Psychiatry, 164, 28-34.
[51] 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
[52] 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
[53] Jon D. Morris. 1995. Observations: SAM: the Self-Assessment Manikin; an efficient cross-cultural measurement of emotional response. Journal of advertising research, 35(6), pp.63-68
[54] 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
[55] 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
[56] R.M. Baños, E. Etchemendy, D. Castilla, A. García-Palacios, S. Quero, and C. Botella. 2012. Positive mood induction procedures for virtual environments designed for elderly people. Interacting with Computers 24, 3 (2012), 131–138. DOI:http://dx.doi.org/10.1016/j.intcom.2012.04.002
[57] Hunter Osking and John A. Doucette. 2019. Enhancing emotional effectiveness of virtual-reality experiences with voice control interfaces. Communications in Computer and Information Science (2019), 199–209. DOI:http://dx.doi.org/10.1007/978-3-030-23089-0_15
[58] 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 (2011), 99–105. DOI:http://dx.doi.org/10.1089/cyber.2010.0100\
[59] Andreas Mühlberger, Roland Neumann, Ljubica Lozo, Mathias Müller, and Marc Hettinger. 2012. Bottom-up and top-down influences of beliefs on emotional responses: fear of heights in a virtual environment. Stud. Health Technol. Inform. 181, 133–137.
[59] American Psychological Association (APA). 2002. Ethical principles of psychologists and code of conduct. Available at: http://www.apa.org/ethics/code/index.aspx.
[60] K. M. Stanney and W. Sadowski. 2002. Presence in virtual environments. In K. M. Stanney (Ed.), Handbook of virtual Environments. Design, Implementation, and applications (p.791-806). Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.
[61] Stephen D. White. 1998. The politics of anger. In B. Rosenwein (Ed.), Anger's past (pp. 127-152). Ithaca, NY: Cornell University Press.Patricia S. Abril and Robert Plant, 2007. The patent holder's dilemma: Buy, sell, or troll? *Commun. ACM* 50, 1 (Jan, 2007), 36-44. DOI: https://doi.org/10.1145/1188913.1188915.

SUPPLEMENTARY INFORMATION

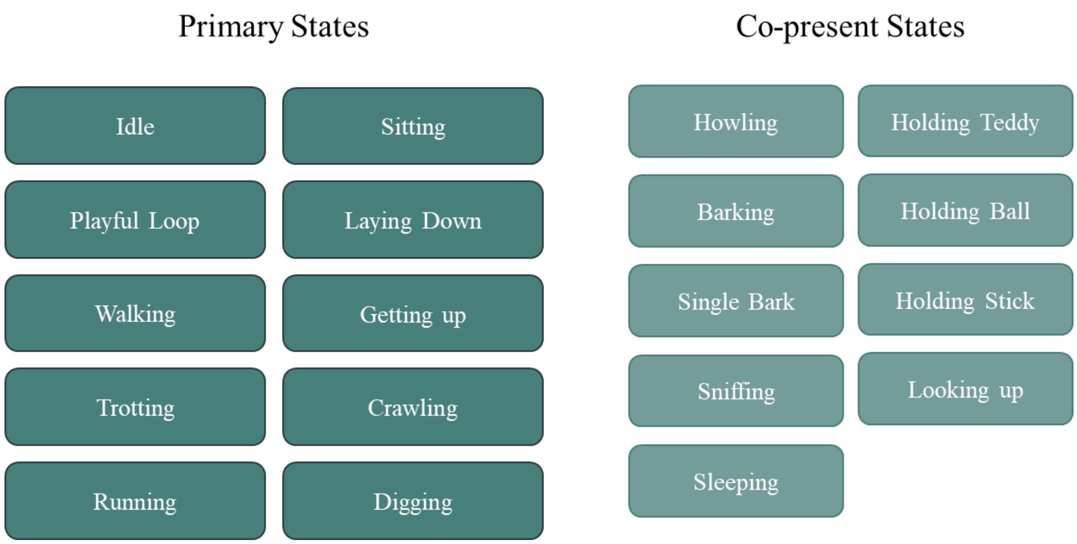
Appendix 1: Flow Diagram of VE Narrative



Appendix 2: State Diagram of Dog's Animations



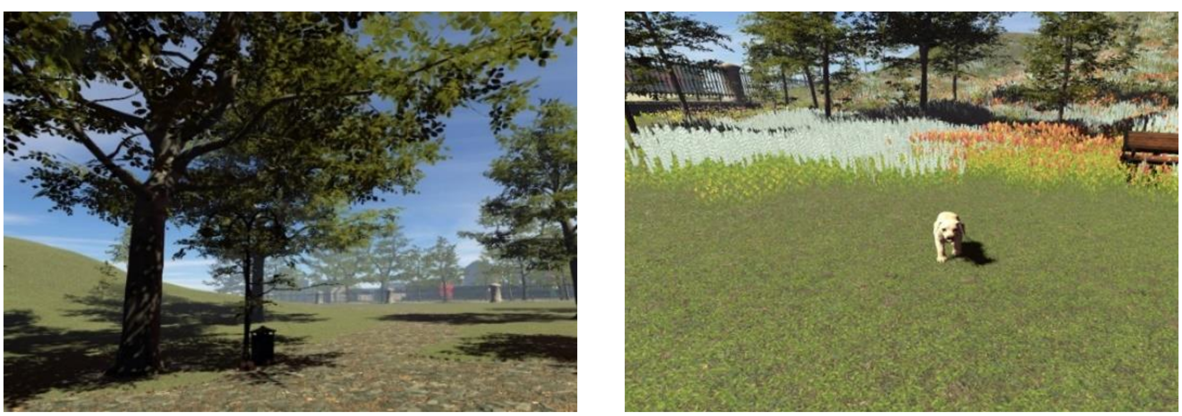
Appendix 3: Dog Animation States



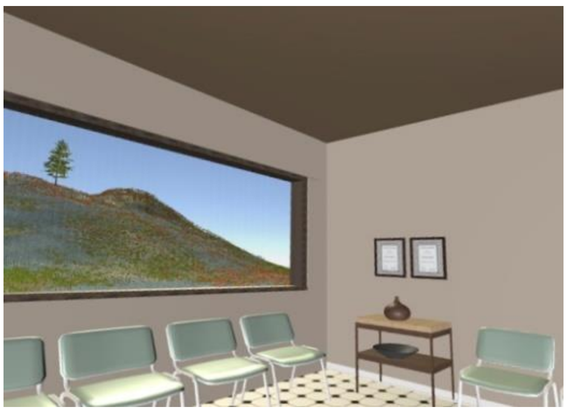
Appendix 4: Updated VE

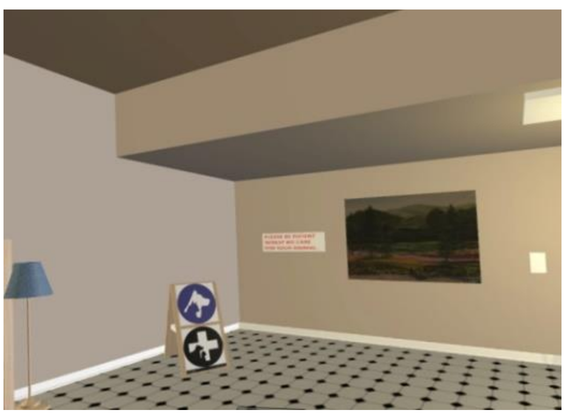
The pet store

The park



The vet





Appendix 5: Pilot Test Notes

Field notes

Participant	Notes
1	• Hand selection = difficult • Fix puppy door • Vet is creepy • Interact with puppy in first scene • Park scene beginning = black
2	• Not sure what to do in first scene • Fix puppy door • Introduce dog • Faulty park scene beginning • Dog must play closer • Make death clearer • Teleported into the road when the puppy ran out • Teleported into the grove and got lost • Headset was blurry (dirty) • What must I do at the vet? • What happened to the dog?
3	• Tutorial = confusing – practice? • Fix puppy door • Fix beginning of park • Hand glitching when close to dog • Female hands disappear • Hover doesn't happen for hand style • Headset glitching • Wanted to pick up teddy
4	• Dog must not run away immediately • Fix puppy door • Teleported to the dog pen – it ended too abruptly • Park beginning must be fixed • Grass looks like it's growing every few seconds
5	• Lagging • Fix puppy door • Park beginning • Got caught in grove teleport barrier • Stick disappeared • Teddy must be pick-up-able • Add a yelp when dog is hit • Dog was barking the entire time – fix • Make it clear you can't touch the dog • Chairs in vet = very big • Vet = scary, looks aimless • Make pup more energetic
6	• Hands = confusing, female hands disappeared • Snap left instruction is never disabled • Fix puppy door • Black at start of park • Make teddy pick-up-able • Threw ball into grove barrier • Vet scene + park = jittery • Make vet hold something • Dog = v masculine, maybe make it softer, add fur?

Code name	Description	Quote	Number of users

Tutorial confusion	Hand selection was the most difficult part of the tutorial. Users also wanted more practice and clearer instructions.	“How do I select this?”	3
Pet store confusion	Most users were not sure what to do.	“Do I go up to the dog?”	2
Park confusion	The dog’s death was unclear, and the teddy scene was confusing for users who teleported into the grove.	“What happened to the dog?”	2
Vet confusion	Users were unsure of what to do, and what the vet’s purpose was.	“What must I do at the vet?”	2
Presence disruption in pet store	The pen door bugs out and does not open correctly.	N/A	6
Presence disruption in park	Grass looks like it is growing every few seconds because of the wind settings, not like it is just blowing in the wind. Some users also threw the tennis ball into the barriers, meaning it randomly dropped out of the sky for an unknown reason.	“Is the grass regrowing every few seconds?”	3
Presence disruption in vet	The vet is almost like a jump scare.	“WOAH that was creepy”	2
Bonding disruption in pet store	The puppy was not really introduced, and most users felt that they had no chance to interact with the puppy. Some users also teleported directly to the puppy pen, meaning the scene ended too quickly for them to get a sense of what was happening.	N/A	3
Bonding disruption in park	Overall, users felt that the dog should play closer to them and be more cute/energetic. The dog sounds were also timed incorrectly.	“Where is she going?”	3
Interaction failures	Most users wanted to pick up the teddy and felt that it was not clear enough that they would not receive feedback when stroking the dog in level 1. Their hands were also jittery when stroking the dog.	“Am I stroking her?”	4
Technical problems in general	Users encountered lagging and the headset being blurry.	N/A	4
Technical problems in tutorial	Snap left instruction not disabled; female hand selection does not work.	N/A	2
Technical problems in park	The beginning of the scene was black until the users teleported. The stick disappeared once dropped by the dog. The users were able to teleport into the road after the puppy died, and some did.	“Where am I?”	6

Appendix 6: Exclusion Criteria

State-Trait Anxiety Inventory (STAI)

The STAI indicates whether participants suffer from general symptoms of anxiety [38]. It is well documented in the literature as both a testing measure [7] and exclusion criterion [8]. It is a self-report questionnaire comprising 20-questions measured on a 4-point scale [38]. Participants with an STAI score greater than 45 are classified as having high general anxiety [39] and were excluded from this study.

Beck Depression Inventory – II (BDI-II)

The BDI-II measures the cognitive aspects of depression in individuals [35]. It is one of the most widely used instruments for identifying depressive symptoms [36] because of its history of successful diagnoses and adequate retest reliability [37]. Participants with a BDI-II score of greater than or equal to 21 likely suffer from moderate to severe depression, and were excluded from this study.

Primary Care PTSD Screen for DSM-5 (PC-PTSD-5)

This screening tool identifies people who likely have PTSD. If participants answer "yes" to three of the five questions, it was assumed that they could have PTSD and were excluded from the experiment (ISTSS, 2015).

Appendix 7: Differences Between Baseline Measures

Variable	Differences between Out-of- and In-Immersion		
	BO	BI	P-value (One-tailed)
Physiological measures M (SD)			
Level 1 B-P			
HR	4.270015	5.035945	0.331294
RR	-2.995840	-1.087171	0.044295
RSA	5.958477	4.424792	0.425136
SCL	0.156089	0.063672	0.338989
Level 1 B-V			
HR	2.934922	3.700851	0.334601
RR	-1.962231	-1.693715	0.423344
RSA	10.720393	4.964109	0.265209
SCL	0.234676	0.076046	0.355266
Level 2 B-P			
HR	3.900027	4.502207	0.316596
RR	-2.261006	-1.986683	0.355312
RSA	1.354780	-1.170002	0.269394
SCL	0.653179	0.530481	0.317454
Level 2 B-V			
HR	6.865537	7.467717	0.415959
RR	-1.342266	-1.067943	0.366686
RSA	-1.040981	-3.565764	0.342781
SCL	0.997722	0.855127	0.327995

Appendix 8: User Study Procedure:

Set-up (0 mins):

1. Export and clear data from VU-AMS SD card
2. Connect VU-AMS device to laptop and enter test parameters

Pre-immersion activities (10 mins)

3. Sign consent form, check if there are any questions
4. Ask demographic questions (studies, year of study, VR/gaming experience)
5. Ask for consent to have electrodes applied
6. Apply electrodes and measure parameter input (back length)
7. Attach electrode wires and SCL electrodes
8. Start VU-AMS recording
9. Put on Headset and attach hand controllers, ensuring that hands are facing the right way
10. Explain controllers and environment

During immersion (6 mins)

11. Tutorial
12. Pet Store
13. Park
14. Vet

Post immersion (10 mins)

15. VAS
16. SAM
17. DEQ
18. Presence
19. Debriefing
20. Payment and Proof of Payment

Appendix 9: Descriptive Statistics and Summaries of Tests

* Indicates data that is not normally distributed

VAS

L1 VAS Sadness		*L2 VAS Sadness*	
Mean	7.653846	Mean	6.75
Median	8.5	Median	7.5
Mode	10	Mode	8
Standard Deviation	2.267383	Standard Deviation	2.562846
Sample Variance	5.141026	Sample Variance	6.568182
Kurtosis	-1.28042	Kurtosis	-0.39086
Skewness	-0.45852	Skewness	-0.67804
Range	6	Range	8
Minimum	4	Minimum	2
Maximum	10	Maximum	10
Count	13	Count	12

t-Test: Two-Sample Assuming Equal Variances

	L1 VAS Sadness n=13	*L2 VAS Sadness* n=12
Mean	7.653846	6.75
Variance	5.141026	6.568182
t Stat	0.935604	
P(T<=t) one-tail	0.179599	
t Critical one-tail	1.713872	
P(T<=t) two-tail	0.359198	
t Critical two-tail	2.068658	

*L1 VAS Surprise**		*L2 VAS Surprise*	
Mean	6.8	Mean	7.875
Median	8	Median	8.5
Mode	4	Mode	9
Standard Deviation	2.588436	Standard Deviation	2.03101
Sample Variance	6.7	Sample Variance	4.125
Kurtosis	-3.21452	Kurtosis	-1.24864
Skewness	-0.50166	Skewness	-0.59468
Range	5	Range	5
Minimum	4	Minimum	5
Maximum	9	Maximum	10
Count	5	Count	8

Mann-Whitney U test

	L1 VAS Surprise n=21	*L2 VAS Surprise* n=22
Mean	7.875	6.8
Variance	4.125	6.7
U-value	13.5	
U critical	6	
Z-score	0.87831	
P-value	0.37886	

SAM

L1 SAM Row 1		*L2 SAM Row 1**	
Mean	3.761905	Mean	3.772727
Median	4	Median	4
Mode	4	Mode	4
Standard Deviation	0.889087	Standard Deviation	0.972567
Sample Variance	0.790476	Sample Variance	0.945887
Kurtosis	0.455646	Kurtosis	2.153659
Skewness	-0.8983	Skewness	-1.20702
Range	3	Range	4
Minimum	2	Minimum	1
Maximum	5	Maximum	5
Count	21	Count	22

Mann-Whitney U test

	L1 SAM Row 1 n=21	*L2 SAM Row 1* n=22
Mean	3.761905	3.772727
Variance	0.790476	0.945887
U-value	228.5	
U critical	0.434126	
Z-score	0.04859	
P-value	0.96012	

L1 SAM Row 2		*L2 SAM Row 2*	
Mean	3.285714	Mean	3
Median	3	Median	3
Mode	4	Mode	2
Standard Deviation	1.230563	Standard Deviation	1.069045
Sample Variance	1.514286	Sample Variance	1.142857
Kurtosis	-1.15746	Kurtosis	-1.15132
Skewness	-0.07263	Skewness	0
Range	4	Range	4
Minimum	1	Minimum	1
Maximum	5	Maximum	5
Count	21	Count	22

t-Test: Two-Sample Assuming Equal Variances

	L1 Presence n=21	*L2 Presence* n=22
Mean	3.285714	3
Variance	1.514286	1.142857
t Stat	0.813894	
P(T<=t) one-tail	0.210204	
t Critical one-tail	1.682878	
P(T<=t) two-tail	0.420408	
t Critical two-tail	2.019541	

L1 SAM Row 3		*L2 SAM Row 3*	
Mean	3.571429	Mean	3.636364
Median	3	Median	4
Mode	5	Mode	5
Standard Deviation	1.325573	Standard Deviation	1.328997
Sample Variance	1.757143	Sample Variance	1.766234
Kurtosis	-1.27434	Kurtosis	-0.57608
Skewness	-0.24113	Skewness	-0.72478
Range	4	Range	4
Minimum	1	Minimum	1
Maximum	5	Maximum	5
Count	21	Count	22

t-Test: Two-Sample Assuming Equal Variances

	L1 Presence n=21	*L2 Presence* n=22
Mean	3.571429	3.636364
Variance	1.757143	1.766234
t Stat	-0.16692	
P(T<=t) one-tail	0.434126	
t Critical one-tail	1.682878	
P(T<=t) two-tail	0.868251	
t Critical two-tail	2.019541	

DEQ

L1 DEQ Fear		*L2 DEQ Fear*	
Mean	2.285714	Mean	2.215909
Median	1.25	Median	2
Mode	1	Mode	2.25
Standard Deviation	1.630348	Standard Deviation	1.160602
Sample Variance	2.658036	Sample Variance	1.346997
Kurtosis	0.404965	Kurtosis	1.015323
Skewness	1.183272	Skewness	1.204233
Range	5.25	Range	4.25
Minimum	1	Minimum	1
Maximum	6.25	Maximum	5.25
Count	21	Count	22

t-Test: Two-Sample Assuming Equal Variances

	L1 DEQ Fear n=21	*L2 DEQ Fear* n=22
Mean	2.285714	2.215909
Variance	2.658036	1.346997
t Stat	0.162341	
P(T<=t) one-tail	0.435918	
t Critical one-tail	1.682878	
P(T<=t) two-tail	0.871835	
t Critical two-tail	2.019541	

L1 DEQ Anxiety		*L2 DEQ Anxiety*	
Mean	2.678571	Mean	2.863636
Median	2.5	Median	3
Mode	2	Mode	3
Standard Deviation	1.167644	Standard Deviation	1.202
Sample Variance	1.363393	Sample Variance	1.444805
Kurtosis	-0.40336	Kurtosis	-0.765
Skewness	0.583872	Skewness	0.195087
Range	4	Range	4.25
Minimum	1	Minimum	1
Maximum	5	Maximum	5.25
Count	21	Count	22

t-Test: Two-Sample Assuming Equal Variances

	L1 DEQ Anxiety n=21	*L2 DEQ Anxiety* n=22
Mean	2.678571	2.863636
Variance	1.363393	1.444805
t Stat	-0.51175	
P(T<=t) one-tail	0.305785	
t Critical one-tail	1.682878	
P(T<=t) two-tail	0.611571	
t Critical two-tail	2.019541	

L1 DEQ Desire		*L2 DEQ Desire*	
Mean	2.845238	Mean	3.261364
Median	2.5	Median	2.875
Mode	1.5	Mode	2.75
Standard Deviation	1.370393	Standard Deviation	1.419417
Sample Variance	1.877976	Sample Variance	2.014746
Kurtosis	1.707385	Kurtosis	0.100505
Skewness	1.068313	Skewness	0.636696
Range	5.75	Range	5.75
Minimum	1	Minimum	1
Maximum	6.75	Maximum	6.75
Count	21	Count	22

t-Test: Two-Sample Assuming Equal Variances

	L1 DEQ Desire n=21	*L2 DEQ Desire* n=22
Mean	2.845238	3.261364
Variance	1.877976	2.014746
t Stat	-0.97727	
P(T<=t) one-tail	0.167084	
t Critical one-tail	1.682878	
P(T<=t) two-tail	0.334168	
t Critical two-tail	2.019541	

L1 DEQ Relaxation		*L2 DEQ Relaxation*	
Mean	3.43254	Mean	4.056818
Median	3	Median	4.375
Mode	1	Mode	1.75
Standard Deviation	1.774162	Standard Deviation	1.453509
Sample Variance	3.147652	Sample Variance	2.112689
Kurtosis	-1.2434	Kurtosis	-0.56818
Skewness	0.163155	Skewness	-0.08136
Range	5.5	Range	5.25
Minimum	1	Minimum	1.75
Maximum	6.5	Maximum	7
Count	21	Count	22

t-Test: Two-Sample Assuming Equal Variances

	L1 DEQ Relaxation n=21	*L2 DEQ Relaxation* n=22
Mean	3.43254	4.056818
Variance	3.147652	2.112689
t Stat	-1.26479	
P(T<=t) one-tail	0.106545	
t Critical one-tail	1.682878	
P(T<=t) two-tail	0.213089	
t Critical two-tail	2.019541	

L1 DEQ Sadness		*L2 DEQ Sadness*	
Mean	3	Mean	2.965909
Median	3	Median	2.75
Mode	3	Mode	2.75
Standard Deviation	1.072381	Standard Deviation	1.421132
Sample Variance	1.15	Sample Variance	2.019616
Kurtosis	-0.75882	Kurtosis	1.762333
Skewness	-0.04201	Skewness	1.166497
Range	3.75	Range	6
Minimum	1	Minimum	1
Maximum	4.75	Maximum	7
Count	21	Count	22

t-Test: Two-Sample Assuming Equal Variances

	L1 DEQ Sadness n=21	*L2 DEQ Sadness* n=22
Mean	3	2.965909
Variance	1.15	2.019616
t Stat	0.088469	
P(T<=t) one-tail	0.464968	
t Critical one-tail	1.682878	
P(T<=t) two-tail	0.929935	
t Critical two-tail	2.019541	

L1 DEQ Happiness		*L2 DEQ Happiness*	
Mean	3.904762	Mean	4.522727
Median	4.25	Median	4.875
Mode	5.5	Mode	4.25
Standard Deviation	1.951724	Standard Deviation	1.854706
Sample Variance	3.809226	Sample Variance	3.439935
Kurtosis	-1.20745	Kurtosis	-0.78778
Skewness	-0.26478	Skewness	-0.57127
Range	5.75	Range	6
Minimum	1	Minimum	1
Maximum	6.75	Maximum	7
Count	21	Count	22

t-Test: Two-Sample Assuming Equal Variances

	L1 DEQ Happiness n=21	*L2 DEQ Happiness* n=22
Mean	3.904762	4.522727
Variance	3.809226	3.439935
t Stat	-1.06461	
P(T<=t) one-tail	0.14664	
t Critical one-tail	1.682878	
P(T<=t) two-tail	0.293281	
t Critical two-tail	2.019541	

Presence

L1 Presence		*L2 Presence*	
Mean	6.047619	Mean	6.181818
Median	6	Median	7
Mode	3	Mode	8
Standard Deviation	2.464268	Standard Deviation	2.788349
Sample Variance	6.072619	Sample Variance	7.774892
Kurtosis	-0.9021	Kurtosis	-0.67294
Skewness	0.083635	Skewness	-0.68424
Range	8	Range	9
Minimum	2	Minimum	1
Maximum	10	Maximum	10
Count	21	Count	22

t-Test: Two-Sample Assuming Equal Variances

	L1 Presence n=21	*L2 Presence* n=22
Mean	6.047619	6.181818
Variance	6.072619	7.774892
t Stat	-0.16692	
P(T<=t) one-tail	0.434126	
t Critical one-tail	1.682878	
P(T<=t) two-tail	0.868251	
t Critical two-tail	2.019541	

HR

L1 HR P-BI		*L2 HR P-BI*	
Mean	5.035945	Mean	4.502207
Median	4.573729	Median	3.948592
Mode	#N/A	Mode	#N/A
Standard Deviation	3.365386	Standard Deviation	3.035768
Sample Variance	11.32582	Sample Variance	9.21589
Kurtosis	-0.97775	Kurtosis	0.591799
Skewness	0.1678	Skewness	1.041848
Range	11.26089	Range	11.36004
Minimum	-0.34475	Minimum	0.703511
Maximum	10.91613	Maximum	12.06355
Count	13	Count	21

t-Test: Two-Sample Assuming Unequal Variances

	L1 P-BI n=13	*L2 P-B1* n=21
Mean	5.035945	4.502207
Variance	11.32582	9.21589
t Stat	0.466317	
P(T<=t) one-tail	0.322597	
t Critical one-tail	1.710882	
P(T<=t) two-tail	0.645193	
t Critical two-tail	2.063899	

L1 HR V-BI		*L2 HR V-BI*	
Mean	3.700851	Mean	7.467717
Median	3.122681	Median	4.408505
Mode	#N/A	Mode	#N/A
Standard Deviation	4.330866	Standard Deviation	8.420206
Sample Variance	18.7564	Sample Variance	70.89987
Kurtosis	1.77045	Kurtosis	1.902332
Skewness	0.060726	Skewness	1.399434
Range	18.32724	Range	33.30298
Minimum	-5.42138	Minimum	-2.14247
Maximum	12.90586	Maximum	31.16051
Count	13	Count	21

t-Test: Two-Sample Assuming Unequal Variances

	L1 V-BI n=13	*L2 V-BI* n=21
Mean	3.700851	7.467717
Variance	18.7564	70.89987
t Stat	-1.71594	
P(T<=t) one-tail	0.048076	
t Critical one-tail	1.695519	
P(T<=t) two-tail	0.096152	
t Critical two-tail	2.039513	

RR

L1 RR P-BI		*L2 RR P-BI*	
Mean	-2.76338	Mean	-1.98668
Median	-2.58805	Median	-2.03367
Mode	#N/A	Mode	#N/A
Standard Deviation	2.442499	Standard Deviation	2.077719
Sample Variance	5.965801	Sample Variance	4.316918
Kurtosis	-0.95053	Kurtosis	-0.15414
Skewness	-0.0368	Skewness	-0.10603
Range	8.249709	Range	8.249715
Minimum	-6.96462	Minimum	-6.04247
Maximum	1.285085	Maximum	2.207247
Count	14	Count	20

t-Test: Two-Sample Assuming Unequal Variances

	L1 RR P-BI n=14	*L2 RR P-BI* n=20
Mean	-2.76338	-1.98668
Variance	5.965801	4.316918
t Stat	-0.96938	
P(T<=t) one-tail	0.170821	
t Critical one-tail	1.708141	
P(T<=t) two-tail	0.341642	
t Critical two-tail	2.059539	

L1 RR V-BI		*L2 RR V-BI*	
Mean	-1.93222	Mean	-1.06794
Median	-1.28181	Median	-0.95568
Mode	#N/A	Mode	#N/A
Standard Deviation	3.42483	Standard Deviation	1.825406
Sample Variance	11.72946	Sample Variance	3.332109
Kurtosis	-1.31915	Kurtosis	0.153648
Skewness	-0.39068	Skewness	-0.07264
Range	9.891608	Range	7.52454
Minimum	-7.82433	Minimum	-4.94621
Maximum	2.067276	Maximum	2.578325
Count	14	Count	20

t-Test: Two-Sample Assuming Unequal Variances

	L1 RR V-BI n=14	*L2 RR V-BI* n=20
Mean	-1.93222	-1.06794
Variance	11.72946	3.332109
t Stat	-0.86237	
P(T<=t) one-tail	0.199913	
t Critical one-tail	1.734064	
P(T<=t) two-tail	0.399825	
t Critical two-tail	2.100922	

RSA

L1 RSA P-BI*		L2 RSA P-BI*	
Mean	3.569664	Mean	-1.17
Median	0.860912	Median	-1.02679
Mode	#N/A	Mode	#N/A
Standard Deviation	23.15761	Standard Deviation	11.29505
Sample Variance	536.2751	Sample Variance	127.5782
Kurtosis	2.159924	Kurtosis	2.322087
Skewness	0.618338	Skewness	-0.71658
Range	100.0962	Range	53.22218
Minimum	-40.8944	Minimum	-31.8333
Maximum	59.20175	Maximum	21.38885
Count	14	Count	19

Mann-Whitney U test

	L1 RSA P-BI n=14	L2 RSA P-BI n=19
Mean	3.569664	1.17
Variance	536.2751	127.5782
U-value	118	
U critical	78	
Z-score	-0.52818	
P-value	0.59612	

L1 RSA V-BI		L2 RSA V-BI	
Mean	8.33158	Mean	-3.56576
Median	6.072203	Median	-0.45833
Mode	#N/A	Mode	#N/A
Standard Deviation	21.51336	Standard Deviation	18.43476
Sample Variance	462.8247	Sample Variance	339.8403
Kurtosis	0.063827	Kurtosis	1.623798
Skewness	0.668387	Skewness	0.28381
Range	73.10393	Range	83.43194
Minimum	-20.9952	Minimum	-39.7444
Maximum	52.1087	Maximum	43.6875
Count	14	Count	19

t-Test: Two-Sample Assuming Unequal Variances

	L1 RSA V-BI n=14	L2 RSA V-BI n=19
Mean	8.33158	-3.56576
Variance	462.8247	339.8403
t Stat	1.666856	
P(T<=t) one-tail	0.054012	
t Critical one-tail	1.708141	
P(T<=t) two-tail	0.108024	
t Critical two-tail	2.059539	

SCL

L1 SCL P-BI		L2 SCL P-BI	
Mean	-0.11169	Mean	0.510584
Median	-0.23468	Median	0.513139
Mode	-0.30939	Mode	#N/A
Standard Deviation	0.291007	Standard Deviation	0.743125
Sample Variance	0.084685	Sample Variance	0.552235
Kurtosis	0.069229	Kurtosis	0.898498
Skewness	1.126057	Skewness	0.128973
Range	0.73719	Range	3.354479
Minimum	-0.36286	Minimum	-1.15025
Maximum	0.374332	Maximum	2.204228
Count	6	Count	20

t-Test: Two-Sample Assuming Unequal Variances

	L1 SCL P-BI n=6	L2 SCL P-BI n=20
Mean	-0.11169	0.510584
Variance	0.084685	0.552235
t Stat	-3.04633	
P(T<=t) one-tail	0.002961	
t Critical one-tail	1.717144	
P(T<=t) two-tail	0.005922	
t Critical two-tail	2.073873	

L1 SCL V-BI	
Mean	-0.0331
Median	-0.14739
Mode	-0.57883
Standard Deviation	0.63124
Sample Variance	0.398465
Kurtosis	-0.03245
Skewness	0.909352
Range	1.590187
Minimum	-0.57883
Maximum	1.011357
Count	6

L2 SCL V-BI	
Mean	0.855127
Median	0.738922
Mode	#N/A
Standard Deviation	0.908538
Sample Variance	0.825441
Kurtosis	-0.5527
Skewness	0.371252
Range	3.498904
Minimum	-0.67965
Maximum	2.81925
Count	20

t-Test: Two-Sample Assuming Unequal Variances

	L1 SCL V-BI n=6	*L2 SCL V-BI* n=20
Mean	-0.0331	0.855127
Variance	0.398465	0.825441
t Stat	-2.70677	
P(T<=t) one-tail	0.009535	
t Critical one-tail	1.782288	
P(T<=t) two-tail	0.019069	
t Critical two-tail	2.178813	

Appendix 10: Relationship between Presence and Number of Interactions Performed by User