Playing with Virtual Reality: Eliciting PLAY through a Water Balloon Virtual Environment

Devin Marais
University of Cape Town
Department of Computer Science
Cape Town, South Africa
mrsdev003@myuct.ac.za

Abstract

The study of emotions requires the effective elicitation of emotions in a controlled environment and recently, virtual reality has gained traction as a medium for doing so. Current VR-based emotion elicitation studies have focused on consciously reported emotions, where few attempts have been made at eliciting a primary-process emotion system from Panksepp's Theory of Affective Neuroscience, with no frameworks for designing such an affective VE existing either. We adapt an existing dimensional-focused affective VE design framework to capture a Pankseppian system's innate triggers, suppressors and categorical signature. We then apply said framework to elicit PLAY, the primary-process underlying reciprocal social joy, through a water balloon fight VE with a skills-matched NPC. Through a within-subjects design 25 participants experienced a tutorial VE, as a control for VR's novelty effects, and the affective Water Balloon VE with three points of measurement via various self-report questionnaires and throw telemetry. The VE successfully evoked joy and increased felt playfulness, whilst creating a sense of sociality with the NPC. However, its elicitation of PLAY's dimensional mapping and behavioural reciprocity fell short on criterion. Therefore, we were not able to elicit PLAY, despite success in eliciting some of its constituent components. These findings suggest that it is possible to elicit a Pankseppian primary process through VR, where the proposed adapted design framework would need to be applied on other Pankseppian systems.

Keywords

virtual reality, virtual environments, affective computing, presence, emotion elicitation, PLAY, Panksepp, emotion systems, affective neuroscience

1 Introduction

The study of emotions requires us to be able to elicit emotions in a controlled environment, such that they may be better studied and understood. Emotion elicitation began with passive techniques whereby participants passively experienced the affective (emotion-inducing) content. While these methods saw success, their lack in ecological validity made their findings less applicable to real world scenarios. Active methods of emotion elicitation where participants play an active role in the affective content have been able to evoke emotions with a greater ecological validity. Of these active techniques, VR has grown in popularity as a medium and has been used to successfully elicit joy, sadness, boredom, fear, awe and other emotions [9, 12, 24]. However, eliciting PLAY, one of the primary emotion processes under Jaak Panksepp's Theory of Affective Neuroscience [20], has not been attempted in the current literature. PLAY refers to the emotion system responsible for feelings of social joy and rough-and-tumble-like behaviour [22]. PLAY is most noticeable during a mammal's juvenile years, and is vital for socio-cognitive development [22] where aberrant PLAY has been linked to early onset schizophrenia, an inability to adapt in changing environments and lower cognitive processing capabilities [1, 4]. However, it has been shown that providing positive PLAY experiences is able to remedy the negative effects of the prior aberrant experiences [13]. Therefore, producing a VE that is able to elicit positive PLAY experiences may be used as a therapeutic intervention. Furthermore, individuals who struggle to engage in PLAY activities with others due to conditions such as autism, may be able to experience PLAY in the VE, where they otherwise would have experienced none. This study aims to determine if the PLAY emotional system can be activated using a VE.

CSC Honours Project, Cape Town, South Africa
2026.

2 Background

In this section, we outline the general concepts in the current affective science field, the varying models that exist to describe emotions and how VR can be used to elicit emotions.

2.1 Models of Emotion

Modern affective science recognises two broad theoretical frameworks for describing emotions, being the discrete model and the dimensional model of emotions [29]. The dimensional model argues that emotions are coordinate combinations of how pleasant the feeling is (valence), how activating the feeling is (arousal) and the degree of control the feeling promotes (dominance) [25]. This produces a three-dimensional plane of valence, arousal and dominance, hereafter VAD. The discrete model holds that emotions come in a limited set of distinct categories [28]. The Theory of Affective Neuroscience [19] is a variant of the discrete model, which argues that emotions are generated by ancient circuits in the subcortical brain, shared by all mammals as a result of evolution. These ancient circuits are called primary process emotional systems which, when activated, generate instinctual patterns of behaviour and a raw affective feeling which is experienced as intrinsically good or bad. Panksepp [19] identified seven of these systems: RAGE, FEAR, CARE, LUST, SEEKING, PLAY, PANIC/GRIEF. The capitalisation of these words is purposeful, and denotes the specific primary process system as opposed to the general definition of each word. That is not to say that all human feelings are in these seven categories, but it is saying that all human feelings at the deepest neurological level, are derived from these systems. The space for more complex and

nuanced feelings is made provision for in Panksepp's [19] three-tier hierarchy of the emotional brain. At the lowest level is the primary process system which is where the subcortical systems producing raw affects and instinctual action reside. One above, at the secondary process level, is where classical and operant conditioning occur, whereby previously neutral cues or stimuli are able to trigger emotional responses due to learned associations. Lastly, the tertiary process level involves higher-order neocortical functions such as expectation, reflection and context-specific emotion interpretation. It is the tertiary process level which creates the nuanced human emotions, by attaching context to the raw affect arising from the primary process. For example, stumbling over your words during a live presentation activates the negatively-valenced primary systems FEAR (perceived danger) and PANIC/GRIEF (social loss). However, it is the tertiary process level which will interpret the situation as "I have made a public mistake, people are judging me" which now makes the emotional episode embarrassment. Furthermore, these three levels bidirectionally interact meaning that during one emotional episode they each influence the activation of each other. In the previous example, the primary process' negative raw affect activated the tertiary level to interpret the feeling into embarrassment. This is bottom-up influence, whereby lower levels influence upper levels. Top-down influence, whereby higher levels influence lower levels, exists as well. For example, if you were attacked by a dog when you were young, your secondary system would learn to associate a dog's bark with FEAR. Then, upon hearing a dog's bark your secondary system activates the primary process FEAR with this learned association, which would activate your tertiary process level to calculate where the barking is coming from. From this, we can see that during any one emotional episode the three levels interact with each other allowing for nuanced and varied emotional responses.

2.2 PLAY as a Primary Emotion System

In this study, however, we focused on Panksepp et al.'s [22] PLAY primary emotion system which is responsible for the rough-and-tumble (chasing, wrestling, pinning) type of behaviour we would observe in juvenile mammals. As previously mentioned, all Panksepian emotional systems produce a raw affect and instinctual behaviour. PLAY's raw affect is generalised to social joy, meaning a highly pleasurable feeling within a social context. The type of behaviour exhibited is the fast-paced, rough-and-tumble actions. In order for PLAY to be activated, the interaction must involve a level of reciprocity and role-reversal. This means that the partners in the interaction alternate between attacking and defending. Furthermore, the interaction must be fair in terms of how often each partner assumes the attacking and the defending role. Both partners must at times be winning, and at other times be losing. Panksepp et al. [22] found that if one partner loses too often, they stop engaging in the interaction and their PLAY system is no longer activated. Even though there is an element of attacking and defending, the actions taken must be signalled as non-serious, so that the partners know the attacking is not a dangerous attack, but rather a playful attack. Lastly, PLAY is regarded as a "luxury" emotional system in that it can only be activated in the absence of the negatively-valenced systems. Hence, PLAY cannot occur if FEAR, RAGE or PANIC/GRIEF are being activated at the same time. Therefore, for PLAY to be activated, its own conditions must be met and the triggers for the negatively-valenced systems must be absent.

2.3 VR as an Emotion Elicitation Medium

In recent years, virtual reality (VR) has gained traction as an effective active emotion elicitation tool. The success of VR in eliciting emotions is most commonly attributed to its ability to invoke a strong sense of presence, which is defined as the sense of "being there" [24]. It has been shown that the level of presence felt is correlated with the emotional intensity felt by the subject [5, 24]. VR has been able to elicit a wide array of emotions. Riva et al. [24] found that through virtual parks, they were able to elicit both anxiety and relaxation in their test subjects. Chirico et al. [9] was able to induce the emotion of awe using 3 separate virtual environments (VE's). Felnhofer et al. [12] developed 5 different virtual parks, each aimed at eliciting different emotions, and succeeded with four out of these five parks. The efficacy of VR in emotion elicitation has been well documented in the current literature [29], where the aforementioned examples are but just a few instances of its success. While there is no single, industry standard design methodology for affective VEs, Dozio et al. [11] have put forward a framework for such purposes. This framework begins by studying the underlying mechanism of the target emotion, to derive both the narrative required to elicit the target and the valence, arousal and dominance levels corresponding to the target. From this, the VE is built by first converting the narrative to a semantic element, which defines the overall mood of the VE, and then selecting individual design elements to modulate the overall valence, arousal and dominance levels of the VE such that the mean of all elements' VAD levels in the VE generalise to that of the target emotion. These individual design elements include sensory components (visual, auditory, olfactory and tactile), interactivity (the influence and control the user has over the VE) and dynamism (how the VE moves or changes over time). The sensory components are able to influence all three dimensions, while interactivity only affects dominance and dynamism only modulates arousal. While the VEs developed under this framework were able to elicit happiness, anger, sadness, fear and disgust [10], they rely on the dimensional model of emotions and therefore cannot be applied to a Panksepian emotion system, such as PLAY.

3 VR Design for Eliciting PLAY

In this section we begin by confronting the drawback of the affective VE design framework from Dozio et al. [11] by introducing new components which capture the Panksepian emotion system in its entirety. Subsequently we apply this adapted framework to create a VE aimed at eliciting PLAY.

3.1 Panksepian VE Design Framework

As discussed in Section 2.3, the affective VE design framework developed by Dozio et al. [11] relies on the dimensional model of emotions, which Panksepian emotional systems are not generally defined in. Furthermore, as Dozio et al. [11] pointed out, not all target emotions can be wholly represented by a VAD coordinate

and require additional contextual framing for that emotion to be elicited. This is definitely true for Panksepian emotion systems. While Panksepian emotion systems do produce raw affect that can be described by the dimensional model, the activation of any one specific system requires that the innate triggers for that system's elicitation are met, the triggers for opposing systems are not met and that the circumstance presented in the VE would activate said system [22]. For example, FEAR would produce low valence, high arousal and low dominance raw affect which can be cleanly captured by the dimensional model. However, FEAR requires for there to be some perceived threat present in the VE. Dozio et al.'s [11] framework would fail to capture this necessary condition. Moreover, as told in Section 2.1, we can activate a primary emotion system through a top-down mechanism. Therefore, it is reasonable to say that an affective VE framework for a Panksepian emotion system should design for such top-down elicitation via learned associations. Lastly, considering that VR's success in its emotion elicitation capabilities has been attributed to its ability to induce presence [29], integrating presence optimisation in an affective VE design framework would make sense. It has been found that presence is influenced by experienced levels of immersion and agency. Immersion speaks to how isolated the subjects feel from the outside world in the VR experience [9]. This level of isolation is mostly related to the technology used in the experience [5] which could be fully-immersive (e.g., head-mounted displays) or non-immersive (e.g., computer screens). It was found that greater levels of immersion were associated with greater levels of presence [5]. Agency refers to the subject's ability to influence, control or alter the VE [11]. Examples of agency include subjects being able to control which parts of the VE they explore and having the ability to move objects in the VE. It has been found that greater levels of agency produce stronger levels of presence [9]. From this, we propose an adapted affective VE design framework for Panksepian emotion systems specifically. The framework begins with defining the emotion system in terms of its dimensional coordinates and its unique categorical aspects which differentiate it from the other primary process systems. It then identifies the innate triggers and suppressors of the primary process at hand. To capture both the qualitative aspects and the trigger profile, we include the affective framing component which comes before the semantic element in Dozio et al.'s [11] original framework. After this stage, the original design framework works as defined, with the focus being on the dimensional coordinate.

Much like Dozio et al. [11], we begin by analysing the target and determining its detectable signatures which we will measure in the evaluation phase. We decompose the target system according to the raw affect it is said to evoke, in terms of its dimensional equivalent and its categorical aspects which make it unique and more than just a coordinate on the dimensional model. We also define both the target system and opposing systems' triggers. The necessity for identifying opposing systems' triggers is because the Panksepian systems can suppress each other if activated at the same time. Hence, we must design the triggers for opposing systems out. Subsequently, we determine the affective framing component which specifies the scenario presented in the VE. This step is necessary because Panksepian systems' activation largely depends on the circumstance the person is in. This scenario realises the categorical signature of the system, stages the innate triggers and designs the suppressing systems' triggers out. This framing also stipulates the secondary process level association, integrated to produce and/or amplify the activation of the primary process. The next step is to follow the framework from Dozio et al. [11]. Once finished, a sweep of the VE is performed to find any places where presence may be optimised according to either agency or immersion. The final step is to define the evaluation tools which will be used to detect the dimensional and categorical signatures of the target primary process system. The full adapted framework is shown in Figure 1.

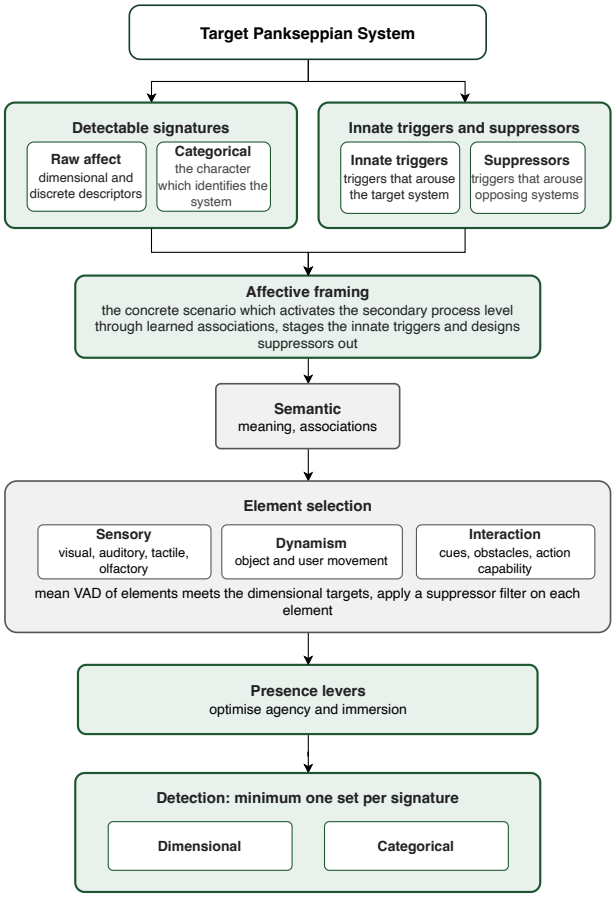


Figure 1: Adapted affective VE design framework from Dozio et al. [11], for Panksepian primary process systems.

3.2 Applying the Adapted Design Framework

In this section, we will now apply the adapted framework from Section 3.1 to build a virtual environment curated to elicit the PLAY system. The completed framework diagram for our PLAY-specific affective VE is shown in Figure 2.

3.2.1 General Description. The affective VE designed for eliciting PLAY placed the participant in a water balloon fight scenario with a young boy non-playable character (NPC) in the backyard of a typical house with surrounding houses visible atop the hedge fences on a bright sunny day. The participant and the NPC are stationed

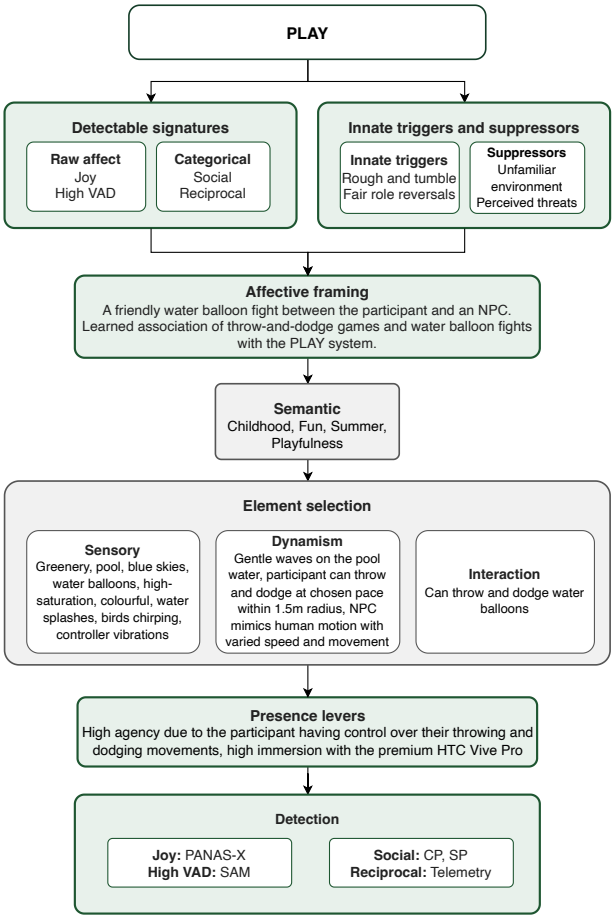


Figure 2: Completed adapted affective VE framework for the Water Balloon VE aimed at eliciting the PLAY system.

diagonally across each other on opposite ends of the pool, where each have buckets of water balloons from which they grab balloons and throw at each other. The VE is designed to ensure the participant experiences guaranteed periods of winning, losing and then fair game-play by measuring the participant's throwing skill level and either increasing, decreasing or matching the NPC's skill level with that of the participant. The NPC also progressively made more advanced moves, including high-velocity throws, two-handed throws, high-throws and side-throws during the experience. There is grass, varied pool toys and a bench visible in the immediate scene. Screenshots of the scene in action are shown in Figure 4 with the overall setting shown in Figure 3.

3.2.2 Detectable Signatures. To determine the dimensional and categorical signatures of PLAY, we must refer to its definition. PLAY is defined as the highly-pleasurable, fast-paced series of back-and-forth interactions between two or more beings, and can be generalised to social joy [22]. From this, we can say that PLAY's raw affect is high valence, high-arousal and high-dominance. The valence and arousal claim come directly from the "highly-pleasurable" and "fast-paced" clauses in the definition, however the dominance claim is



(a) Pool-side props and balloon buckets

Figure 3: Buckets of water balloons, pool toys and a braai placed around the pool where the player starts.

derived from the notion of a series of back-and-forth interactions. Dominance refers to control/influence over one's environment and in the back-and-forth sequence of PLAY, each player's action alters how others can respond and their responses in turn reshape the action space. Therefore, every player continuously influences their environment through their chosen actions in the action space, and as such we say that this constitutes as high dominance. Additionally, the "social joy" generalisation indicates that the raw affect of PLAY can be equated with that of joy. For the categorical signature of PLAY, from the definition's "back-and-forth" description, we can identify a reciprocal nature. Furthermore, the "between entities" assertion indicates it is a social interaction. With this, we say that the categorical signature of PLAY is for there to be both a reciprocal and social element involved.

3.2.3 Innate Triggers and Suppressors. The innate triggers of PLAY are the specific stimuli which activate the PLAY system, where the suppressors are the innate triggers of opposing Panksepian systems to PLAY. According to Panksepp et al. [22], PLAY is reciprocal with role-reversals occurring between the entities involved. Furthermore, in these role-reversals there must be an element of fairness where each entity assumes either role for roughly the same amount of time. In the case of one of the entities being inherently more equipped to assume one of the roles, they must self-handicap themselves such that the other entities may assume this role as well. As an example, in Panksepp's work with rats, he identified pinning to be their main expression of PLAY. In these bouts of PLAY, rats would at some points be pinning another down and at others, be getting pinned down. Here, the two roles are doing the pinning and being pinned, where the rats continuously swapped between the two. The notion of fairness came from how Panksepp observed that if a rat stayed in the role of being pinned down (due to not being strong or nimble enough to pin another rat down) for too long, they would stop playing and their PLAY system would down-regulate, hence why fairness and self-handicapping is necessary to continue PLAY's elicitation. Moreover, in these interactions, it must be signalled that it is non-serious. Lastly, PLAY requires for the person to feel safe, therefore FEAR's trigger of perceived danger must be suppressed. Of these triggers, the main concern would be



(a) NPC releases the balloon **(b) Balloon in flight** **(c) Splash on impact**

Figure 4: The NPC throws a balloon at the player, ending in a splash across the player's view.

any perceived threat or feelings of unsafety due to an unfamiliar surrounding.

3.2.4 Affective Framing. The affective framing realises the learned association with the primary process PLAY and the innate triggers, whilst designing the suppressing triggers out. The VE places the participant in a throw-and-dodge type game with a water balloon fight narrative. Both throw-and-dodge interactions and water balloon fights can be described as generic instances of PLAY, hence they act as the learned associations with PLAY. Regarding PLAY's innate triggers, the VE's reciprocal nature lies in how the participant can both throw and be thrown at by the NPC. Furthermore, the participant and the NPC recurringly switch roles from being on the offence (throwing) to being on the defence (dodging). With that being said, the fairness component is baked in with the NPC skill adjustment where it will adjust its throwing and dodging skills to match that of the participant through modulating its throw frequency, throw accuracy and reaction time in response to the skill of the participant, which is calculated according to those same levers. Moreover, water balloon fights are inherently non-serious as they are not competitive in the way that a soccer match, for example, would be and there are no serious consequences that could arise from the game outcome, such as, for example, making it to a soccer league final. Conversely, by making the NPC a non-threatening young boy and placing the entire scene within a visually clear and defined space (where no hidden threats could lie) the potential of participants perceiving any threats is mitigated. Similarly, by making the scene a generic backyard setting of a typical house, the environment is immediately familiar to the participant which reduces the likelihood of any unsafe feelings.

3.2.5 Semantic and Element Selection. In this phase, the semantic, sensory, dynamism and interaction elements are chosen and rated according to their VAD ratings, which we obtained from the International Affective Picture System [14] and the International Affective Digitised Sounds Database [8]. For our purposes, we have excluded olfactory elements. The full description of these chosen elements is shown in the completed adapted design diagram as shown in Figure 2.

3.2.6 Presence Levers. In this portion of the design process, we determine mechanisms to optimise presence by optimising agency and immersion. In terms of optimising agency, this is achieved already through the highly interactive throwing element of the VE where the user can control what, where and how they throw. Furthermore, they can move their body to enhance their throws or to dodge throws from the NPC. With respect to immersion, this is optimised through the high-quality 2018 HTC Vive Pro headset and controllers we use for delivering the VE experience to participants.

3.2.7 Detection. Here, we must specify how each constituent component from the detectable signatures will be measured, which for our purposes will be through various self-report questionnaires and telemetry. It could be argued that since PLAY is a subcortical system, conscious self-report (which operates cortically) would not accurately capture its elicitation. However, as told in Section 2.1, the elicitation of PLAY at the primary process level will interact with the tertiary process level which is the level that self-report measures capture. Hence, self-report measures will be able to measure PLAY elicitation, albeit indirectly. We will utilise the Self-Assessment Manikin (SAM) [7] to measure experienced levels of valence, arousal and dominance in the participants. Furthermore, we will use the Positive and Negative Affect Scale (PANAS-X) [32] to assess levels of felt joy. Similarly, to evaluate the extent to which the VE felt social to the participant, we will employ the Perceived Other's Co-Presence (CP) [18] and the Social Presence (SP) [18] questionnaires. Contrastingly, to determine if reciprocity was exhibited, telemetry for the participant's throws relative to those of the NPC will be taken.

3.2.8 Tutorial VE. The focus of this study is on the Water Balloon VE's ability to elicit PLAY. There is potential for confound due to the novelty effect of VR whereby first-time VR users experience pleasure and excitement purely due to using VR itself [15]. Conversely, first time VR users may struggle with the controls for interacting with and navigating through the VE which may cause feelings of frustration arising from a VR-skills gap. Both of these cases introduce raw affect not solely attributable to the Water Balloon VE. Therefore, we expose participants to a tutorial VE during which VR novelty effects will be absorbed and participants will become accustomed to the VR controls. During this tutorial VE, participants will be placed in a visually minimalistic scene where they complete tasks for teleportation, snap-turn, grabbing objects and throwing objects. With this in place, we can say that the affective states produced by the Water Balloon VE are a consequence of its specific content. Even though we expect the tutorial environment to be emotionally inductive due to the novelty effect of VR and controller-learning frustration, we will consider the participant's affective state following the tutorial environment, but before the

Water Balloon VE, as their neutral state. We assume this position because the novelty effect of VR and/or controller-frustration may occur in the Water Balloon VE as well, therefore we must compare the affective state following the Water Balloon VE with that which too is influenced by these factors.

4 Experimental Design

In this section, we outline the methodology this study used and the full experimental procedure taken to carry this methodology out. This methodology tests whether PLAY was elicited by the affective VE described in Section 3.2. This study employs a within-subjects design with 25 participants (n = 25), where each participant will first experience a tutorial VE aimed at absorbing any novelty effect from first-time experience of VR and neutralise the emotional state of the participant. Following the tutorial VE, the participant will experience the affective VE. During the experience, the participant will complete three suites of subjective questionnaires which will later be used for statistical analyses to determine if PLAY was indeed elicited. Ethics clearance for this study was granted by the UCT Inter-Faculty Research Ethics Committee, with the reference number IFHREC/02979/2026.

4.1 Participants

25 participants were secured using opportunity recruitment through informal communication channels such as WhatsApp groups for courses at the University of Cape Town (UCT), as well as through direct messages to personal contacts. For our study, no inclusion criteria were in place, however we excluded participants who were below 18 years of age, had existing psychiatric disorders, were on chronic stimulant-medication, had a score of 20 or above on the Beck Depression Inventory (BDI) [6] or had history of motion-sickness. Existing psychiatric disorders influence emotional reactivity [27], which may skew our findings for efficacy of emotion elicitation. Further, the use of stimulant-medications and depression, as indicated with scores of 20 or above on the BDI [6], have been shown to reduce expressions of social play which hinders our VEs ability to elicit PLAY [17, 22]. Cyber-sickness during VR experiences impacts the reported efficacy of affective VEs, and a history of motion-sickness is indicative of a higher probability in experiencing cyber-sickness [23].

4.2 Apparatus

In this study, each participant experienced two VEs, the Tutorial VE and the Water Balloon VE, developed with Unity version 6000.3.8f1 and 6000.3.17f1 respectively. The experiments were conducted in a dedicated VR room in the UCT Computer Science building, using the 2018 HTC Vive Pro headset and two corresponding controllers. The VE was run on a Windows 10 and Unity3D desktop computer with a 2.8 GHz, 6-core Intel® i5 Chip, paired with 16GB of RAM and an NVidia GTX 1070 graphics card. Lastly, the questionnaires detailed in Section 4.3 were administered via paper format and manually captured using Microsoft Excel.

4.3 Measures

As told in Section 3.2.2, we will make use of the SAM, PANAS-X, CP and SP self-report questionnaires along with telemetry of the participant's throws compared to that of the NPC, in determining if PLAY's detectable signatures were observed. The SAM presents 3 sets of 5 images, where each set corresponds to valence, arousal or dominance. The participant selects which image represents their current emotional state the best, using a 1 - 9 scale plotted beneath each image set. The PANAS-X is a multidimensional questionnaire which lists 60 discrete emotions where participants indicate from 1 - 5 how much their current emotional state aligns with each discrete emotion. The 60 items are grouped into 13 different subscales where each subscale represents an overall emotional state such as General Positive Affect, Joviality and Hostility. The CP questionnaire presents participants with 11 statements, which they rate on a 1 -5 agreement scale, regarding how the participant perceived the NPC. Similarly, the SP questionnaire poses 6 questions on a 1 - 7 agreement scale which aim to determine to what extent the participant felt social with the NPC. Lastly, the telemetry recorded includes timestamped throws made by the participant or the NPC and if said throw was successful in hitting the opponent. However, we employed 2 additional questionnaires, being the Affective Neuroscience State Scale (ANSS) [3] and the iGroup Presence Questionnaire (iPQ) [26]. The ANSS is a multidimensional questionnaire, much like the PANAS-X, comprising 40 items that map to Panksepp's 7 primary systems, where PANIC and GRIEF have been separated into 2 separate subscales, resulting in the ANSS having 8 subscales overall. We used this questionnaire to determine which of Panksepp's systems the participant's emotional state related to the most. However, we omitted the LUST subscale items from the ANSS for our purposes following Panksepp's own advise where he indicated participants were likely to be deceitful when answering the LUST subscale items and that this deceit would carry over into subsequent, unrelated questions, therefore causing inaccurate findings. Moreover, since presence to a great extent impacts the emotionally-inductive capability of any VE [24], it would make sense to record the extent to which participants felt present in the VE. We do this using the iPQ, a questionnaire with 14 statement/question items which participants score on a -3 to +3 Likert scale based on their respective agreement/answer with each item. Operationally, not all of these measures will be taken at the same juncture, which in this experiment can be before the Tutorial VE (pre-tutorial/P1), after the Tutorial VE (post-tutorial/P2) or after the Water Balloon VE (post-affective/P3). The SAM, PANAS-X and ANSS will be taken at each time period as a repeated monitoring of participants' emotional states. The iPQ's items relate to a VE experience, and can therefore only be completed at P2 and P3. Similarly, the CP and SP questionnaires relate to an interaction with another social being, which in our case is with the NPC, hence these questionnaires can only be utilised at P3. Lastly, the telemetry will only be recorded while the participant experiences the Water Balloon VE. For the purposes of Section 4.4, we will define 3 suites of measures. The first being the pre-tutorial suite (SAM, PANAS-X, ANSS), the second the post-tutorial suite (SAM, PANAS-X, ANSS, iPQ) and the third as the post-affective suite (SAM, PANAS-X, ANSS, iPQ, CP, SP, telemetry).

4.4 Procedure

Participants complete the experiment one at a time over a duration of approximately 50 minutes. During each session, the participant

was first brought into the dedicated VR room where the researcher introduced themself and gave participants the opportunity to ask any questions (4 min). Following this, participants completed the consent form (2 min) and the pre-tutorial suite of measures (8 min). Subsequently, the participant completed the Tutorial VE (6 min) and, once finished, filled in the post-tutorial suite of measures (9 min). At this stage, the participant experienced the Water Balloon VE (7.5 min) and answered the post-affective suite of measures (11 min), where the telemetry was being recorded throughout the 7.5 min Water Balloon VE experience. Following this, to close off the experiment, the participant was debriefed on what the study was about and was given the opportunity to ask any questions (5 min).

4.5 Data Analysis

Functionally, each of the measures from Section 4.3 have associated hypotheses. The SAM will be used to determine if the participant experienced high levels of valence, arousal and dominance (H-SAM-1v/a/d) where "high" refers to being in the upper tertile of the SAM scale which is any score above 6.33, where this tertile-classification method has been followed in prior work [16]. We also aim to determine if the participant's valence, arousal and dominance levels increased from P2 to P3 (H-SAM-2v/a/d). Furthermore, we will determine if joy was felt through the Joviality subscale of the PANAS-X which roughly corresponds to joy. More specifically, we will test if the change in Joviality from P2 to P3 was greater than that of all other subscales over the same period (H-PANASX-1) and if Joviality has the highest score across all other subscales, at P3 (H-PANASX-2). The ANSS will assess the raw affect's alignment with each Pankseppian system, and for our purposes we want to ascertain if the PLAY subscale increased more than all other subscales, from P2 to P3 (H-ANSS-1) as well as if at P3 the PLAY subscale has the highest mean score across all other subscales (H-ANSS-2). Regarding evaluating the extent to which participants felt the experience was social, we will check if the NPC was perceived as a social being more than not by testing if the mean scores on the CP and SP questionnaires were above their midpoints of 3 (H-CP-1) and 4 (H-SP-1) respectively. To determine if reciprocity was exhibited, we test if participant throws were part of their independent throw rhythm or if they were temporally aligned with NPC throws (H-TEL-1). Using the telemetry, we calculate the proportion of NPC throws answered, meaning the participant threw at least once before the NPC's next throw. Some answered throws may occur by chance under any throwing rhythm, so we estimate this chance using time-shifted surrogate data (Appendix B) which is subtracted from the observed data. The resulting difference would be zero if participants threw on their own rhythm, hence we test if these differences are zero. Lastly, for the presence explanatory variable we will determine if the participants felt present more than not present in the Water Balloon VE through seeing if the mean iPQ score is above the scale midpoint of 0 (H-iPQ-1) and if greater presence was experienced in the Water Balloon VE than the Tutorial VE (H-iPQ-2). While these hypotheses each imply a predicted direction, we will apply two-sided statistical tests due to the possibility of the Water Balloon VE having significance in the opposite predicted direction.

In this study, we make use of three distinct families of tests. The first being a one-sample test against some fixed value (F1), the second being a paired-sample using the same measure for one participant across different readings (F2), and the third being a paired-sample test for different measures from one participant at the same reading (F3). With that being said, we will apply an α of 5% to all tests where each dataset for each test will undergo a test for normality using the Shapiro-Wilk test with an α of 5% and a visual inspection of the associated Q-Q plot. If normality is found, the parametric test will be carried out, and if it is not found, the non-parametric equivalent will be applied and any necessary Holm-Bonferroni correction will be employed in either case. To claim the successful elicitation of PLAY, all stipulated hypotheses must show significance at the 5% level.

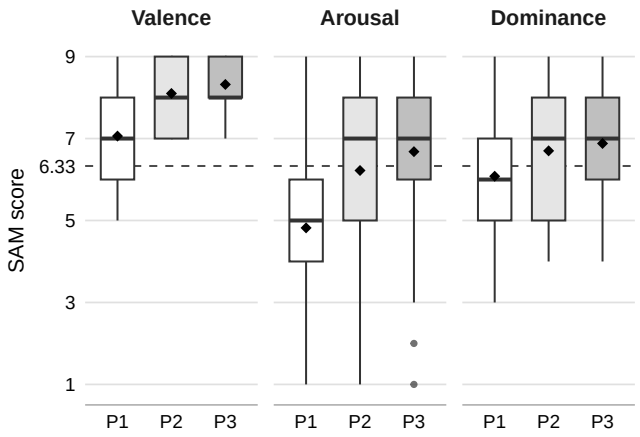


Figure 5: Box and whisker plots for valence, arousal and dominance across all three readings.

5 Results

Here, we stipulate the outcome of each hypothesis from Section 4.5 with visual representations of the data where applicable. Since each hypothesis corresponds to one measure (questionnaire or telemetry) we will state the findings according to this grouping. The full results and relevant statistics can be found in Appendix A.

5.1 SAM

The overall results for the SAM across the three different readings appear in Figure 5, where we can see that the valence, arousal and dominance levels are descriptively above the "high" 6.33 threshold and all increased from P2 to P3. However, only the valence reading at P3 showed significance above the threshold, and no readings showed a significant increase from P2 to P3.

5.1.1 H-SAM-1. The valence, arousal and dominance scores at P3 were not found to be normal and as such, the one-sample Wilcoxon signed-rank test was applied. It was found that the valence at P3 was significantly higher than the high threshold of 6.33 ($V = 325$, $p_{\text{Holm}} < 0.001$), while significance for arousal ($V = 206$, $p_{\text{Holm}} = 0.245$) and dominance ($V = 230$, $p_{\text{Holm}} = 0.141$) was not found. Hence, H-SAM-1v was supported whilst H-SAM-1a and H-SAM-1d were not supported.

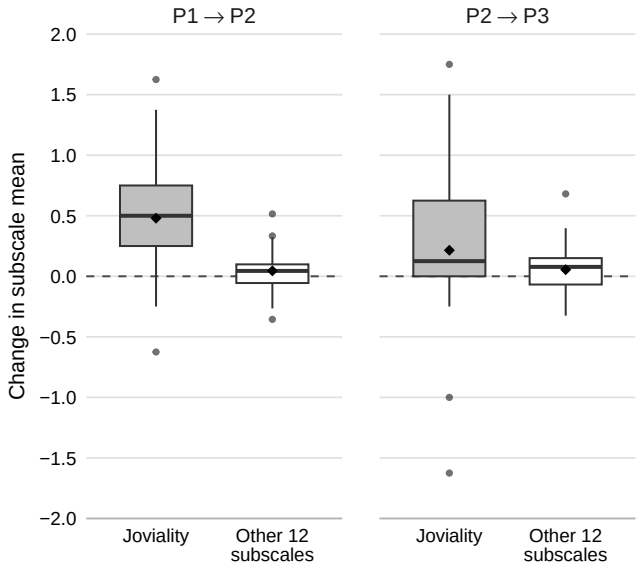


Figure 6: Box and whisker plots for the change in Joviality against that of all other subscales.

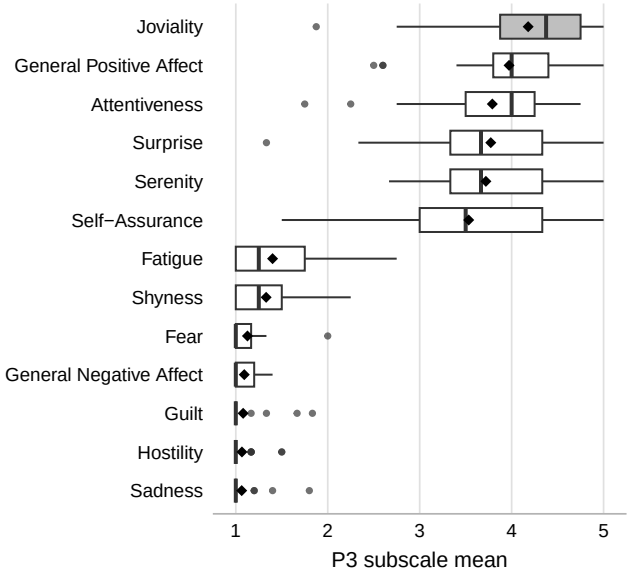


Figure 7: Box and whisker plots for all PANAS-X subscales at P3.

5.1.2 H-SAM-2. The change in valence, arousal and dominance scores from P2 to P3 were not found to be normal, so paired Wilcoxon signed-rank tests were employed. No significant changes from P2 to P3 for valence ($V = 35.0$, $p_{\text{Holm}} = 0.134$), arousal ($V = 92.0$, $p_{\text{Holm}} = 0.217$) or dominance ($V = 63.0$, $p_{\text{Holm}} = 0.517$) were found. Therefore, H-SAM-2v, H-SAM-2a and H-SAM-2d were not supported.

5.2 PANAS-X

From the PANAS-X tests, we found that the Joviality subscale did increase the most from P2 to P3 and had the highest mean item score at P3, compared to the remaining subscales over the same time periods, as shown in Figures 6 and 7.

5.2.1 H-PANASX-1. For this test, the change in Joviality from P2 to P3 was compared with the mean change of the other 12 other subscales. This data was not found to be normal, hence the paired Wilcoxon signed-rank test was performed. From this, a significant difference between the two changes was found, where the change in Joviality was significantly greater ($M_{\text{diff}} = 0.16$, $V = 236$, $p = 0.04951$) than that of the other 12 subscales, as shown in Figure 6. Therefore, H-PANASX-1 is supported.

5.2.2 H-PANASX-2. The difference between the Joviality mean item score and that of each individual other subscale at P3 was only found to be normal for the General Positive Affect, Serenity and Surprise subscales. Hence, t-tests were performed for these three subscales, and Wilcoxon signed-rank tests were carried out for the non-normal, remaining subscales, all paired with Joviality. Holm-Bonferroni correction was applied across all 12 tests, regardless of their normality. It was found that the Joviality subscale was significantly higher than every other subscale, as shown in Figure 7, where the closest runner-up was General Positive Affect ($M_{\text{Joviality}} =$ 4.18, $M_{\text{GeneralPositiveAffect}} = 3.97$, $t = 2.15$, $p_{\text{Holm}} = 0.042$). Therefore, H-PANASX-2 is supported.

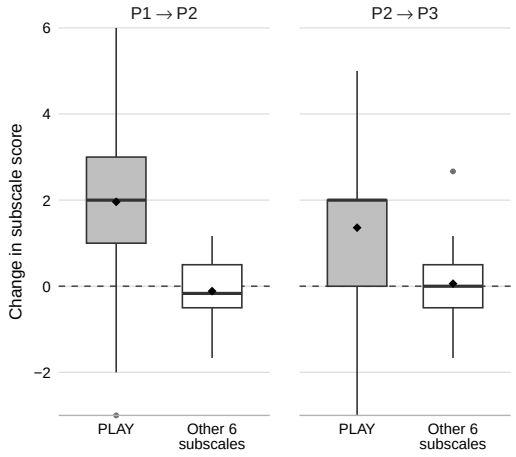


Figure 8: Box and whisker plots for the change in the PLAY subscale against that of all other subscales.

5.3 ANSS

The results from the ANSS tests indicated that the PLAY subscale increased from P2 to P3 significantly more than all other subscales, however at P3, while descriptively it was, PLAY was not statistically found to be significantly higher than all the remaining subscales.

5.3.1 H-ANSS-1. Much like in H-PANASX-1, the change in the PLAY subscale from P2 to P3 was compared with the mean change of the remaining 6 subscales. This data was found to follow a normal distribution, and as such, a paired t-test was conducted with

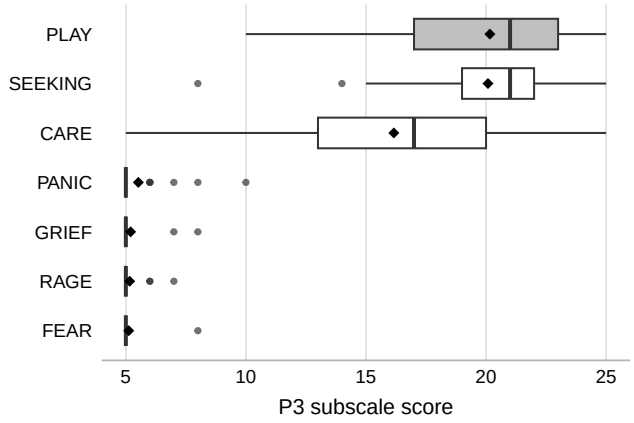


Figure 9: Box and whisker plots for all ANSS subscales at P3.

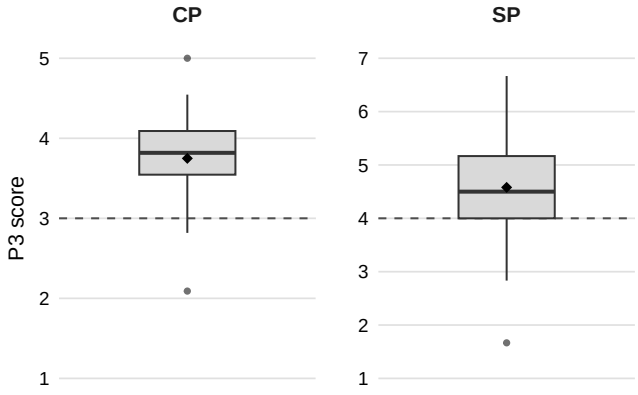


Figure 10: Box and whisker plots showing the Perceived Other's Co-Presence (CP, scale 1 - 5, midpoint 3) and Social Presence (SP, scale 1 - 7, midpoint 4) scores at P3.

the outcome being that the change in the PLAY subscale was significantly greater than that of the other 6 subscales ($M_{\text{diff}} = 1.30$, $t = 3.26$, $p = 0.003$) as depicted in Figure 8. Hence, H-ANSS-1 is supported.

5.3.2 H-ANSS-2. Of the differences between the PLAY subscale score and that of each other subscale, none were found to follow a normal distribution, hence the paired Wilcoxon signed-rank tests were performed. From these tests, the PLAY subscale was significantly higher than all other subscales except for SEEKING ($M_{\text{PLAY}} = 20.16$, $M_{\text{SEEKING}} = 20.08$, $V = 84.5$, $p_{\text{Holm}} = 0.718$) where the next runner up was CARE ($M_{\text{PLAY}} = 20.16$, $M_{\text{CARE}} = 16.16$, $V = 291.5$, $p_{\text{Holm}} = 0.001$) as shown in Figure 9. Hence, H-ANSS-2 is not supported.

5.4 CP and SP

The CP and SP questionnaires were utilised to assess the extent to which the Water Balloon VE felt social. Significance supporting the claim that the VE was experienced as social for both the CP and the SP was found.

5.4.1 H-CP-1. The mean CP score at P3 did not follow a normal distribution, and as such, the one-sample Wilcoxon signed-rank test was applied. It was found that the CP score at P3 was significantly above the midpoint of 3 ($M_{\text{CP}} = 3.75$, $V = 306.5$, $p < 0.001$) as shown in Figure 10. Hence, H-CP-1 is supported.

5.4.2 H-SP-1. The mean SP score at P3 did follow a normal distribution and the one-sample t-test was employed. From this, the mean SP score was found to be significantly higher than the midpoint of 4 ($M_{\text{SP}} = 4.58$, $t = 2.49$, $p = 0.020$) which can be seen in Figure 10. Therefore, H-SP-1 is supported.

5.5 Telemetry

5.5.1 H-TEL-1. The data was found to follow a normal distribution, and a single t-test against the value 0 was performed. Significance was not found ($M_{\text{observed}} = 0.825$, $M_{\text{surrogate}} = 0.816$, $t = 1.316$, $p = 0.202$) and as such, H-TEL-1 is not supported. Figure 11 shows how often the participant threw relative to an NPC throw (at lag = 0) compared to what would be expected if their throws were made independent of the NPC throw, as shown by the surrogate line.

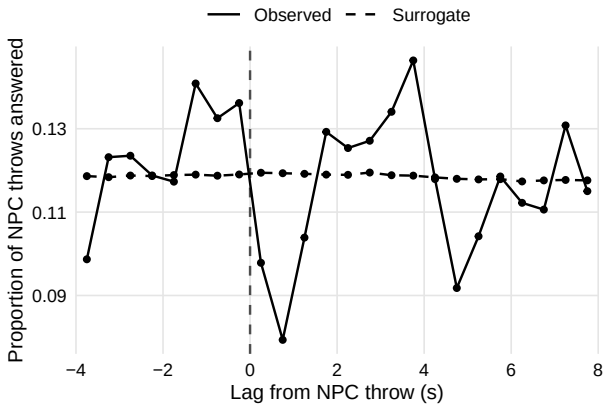


Figure 11: Graph showing the proportion of NPC throws answered relative to the time at which an NPC throw was made (at lag = 0).

5.6 iPQ

Overall, presence was found to be significantly higher than the midpoint of 0 indicating that participants felt present more than not present in the Water Balloon VE. Furthermore, the P3 iPQ score was determined to be significantly higher than that of the P2 IPQ score indicating that the Water Balloon VE induced greater feelings of presence than the Tutorial VE.

5.6.1 H-iPQ-1. The iPQ scores at P3 followed a normal distribution, hence a one-sample t-test was utilised and the iPQ scores were found to be significantly higher than the midpoint of 0 ($M = 1.41$, $t = 10.20$, $p < 0.001$), as shown in Figure 12. Hence, H-iPQ-1 is supported.

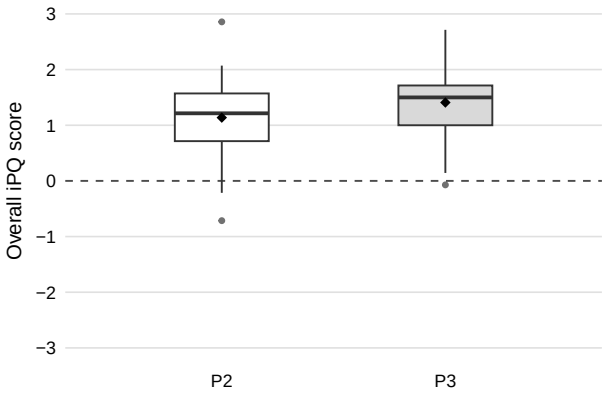


Figure 12: Box and whisker plots showing the P2 and P3 overall iPQ scores (scale from -3 to +3, midpoint of 0).

5.6.2 H-iPQ-2. The iPQ score change from P2 to P3 followed a normal distribution, and a paired t-test was carried out where results indicated that the iPQ score increased significantly from P2 to P3 ($M_{P3} = 1.41$, $M_{P2} = 1.14$, $t = 2.46$, $p = 0.021$) which is depicted in Figure 12. Therefore, H-iPQ-2 is supported.

6 Discussion

In this study, we set out to produce the raw affect and behaviour observed with PLAY's activation within a felt social context. We successfully elicited raw affect aligning with joy, high valence and the discrete descriptors of PLAY within an interaction that was perceived as social. However, we failed to evoke both high arousal and high dominance, as well as PLAY's reciprocal behaviour. Moreover, while Joviality outweighed its sibling subscales, the PLAY subscale did not for its corresponding Panksepian subscales. From these results and our requirements for claiming successful PLAY elicitation, we were unable to elicit PLAY. As mentioned in Section 3.2.8, P2 acted as our control even though we expected for its readings to be influenced by the novelty effect of VR. This may provide insight to why valence did not significantly increase from P2 to P3, as the increase in valence from P2's novelty effect [15] effectively pushed valence to its ceiling ($M = 8.10$). Contrastingly, the failure for arousal and dominance lies in the VE's under-assisted throwing mechanism resulting in an average throw accuracy of 12.6%. With this, participants experienced the same outcome of not hitting the NPC rather often, and this repetition could have resulted in boredom, which in turn lowers arousal [30]. Furthermore, repeated failure likely reduced the extent to which participants felt in control, hence lowering dominance. It is plausible then to say that arousal and dominance were dampened by the throwing failure, and could not increase much beyond what the novelty effect raised the levels to. Additionally, the failure of the PLAY subscale scoring significantly higher than SEEKING, could be attributed to the spike in SEEKING brought about by the novelty effect. The SEEKING system revolves around exploration and curiosity [2, 19], which are prominent outcomes from the novelty effect [15]. From this, we would expect SEEKING to have been elicited to a greater extent than its other subscales at P2 (where the novelty effect was most prominent). In a follow up test, SEEKING increased the most from P1 to P2 and scored the highest across the other subscales. With this, SEEKING is arguably sensitive to the novelty effect which means its very high mean at P3 could be related to the novelty effect, and not due to the Water Balloon VE. Conversely, Panksepp identified SEEKING as a general appetitive system which participates in and energises all other primary process systems [2, 21]. Hence, the result showing both SEEKING and PLAY to be high in scores is expected and not indicative of a failure to elicit PLAY. Moreover, while the claim that participant throws were temporally aligned with that of the NPC did not find significance, in Figure 11, showing how often a participant threw relative to an NPC throw, we see a pattern of throw rates spiking and then dipping - most noticeably a spike leading up to the NPC throw, with a subsequent dip after. This pattern implies a deliberate pause following an NPC throw, with a responding throw being made shortly after which implies a pattern of participant throw, NPC throw, participant throw, NPC throw and so on, therefore suggestive of a reciprocal throwing nature. Lastly, comparing our iPQ scores against the scale midpoint of 0 is arguably weak as a means to evaluate the Water Balloon VE's presence. We can turn to the iPQ score distributions for head-mounted display VR studies developed by Tran et al. [31], and rate our observed mean iPQ score according to it. From this, our P3 iPQ score ($M = 1.409$, $SD = 0.691$, $CI = [1.124, 1.694]$) sits in the high to exceptional range [31]. Hence, we can say that the Water Balloon VE produced high levels of presence. Regarding future work, we suggest the study be carried out with frequent VR-users such that the efficacy of the VE may be better identified in the absence of VR's novelty effects. Furthermore, if the same VE were to be used, improving the throwing mechanism by providing greater participant throw assists is recommended such that arousal and dominance are not reduced due to repeated unsuccessful throws.

7 Conclusion

The use of VR in emotion elicitation has grown dramatically in recent years, yet no attempt at eliciting PLAY or developing an affective VE design framework for Panksepian systems had been made. In this study, we provide both. Through an adaptation of an existing dimensional-based design framework that captures a Panksepian system's innate triggers, suppressors and categorical signature we designed a water balloon fight VE aimed at eliciting PLAY. Against the pre-specified criterion, the VE did not elicit PLAY in full as arousal, dominance and behavioural reciprocity fell short of significance. Nonetheless, the VE evoked a highly-pleasant raw affect, with the Joviality and PLAY subscales increasing most, in an environment that was perceived as social. With the reusable, adapted framework these findings suggest that top-down elicitation for Panksepian systems through VR is partially achievable. The Water Balloon VE's throwing mechanism should be corrected and then tested with experienced VR users.

8 Generative AI Declaration

Generative AI Tools (Claude) were used in the proofreading process for this paper, and as assistance in generating statistical analysis scripts.

References

[1] Ralph Adolphs. 2010. Conceptual Challenges and Directions for Social Neuroscience. *Neuron* 65, 6 (2010), 752–767. doi:10.1016/j.neuron.2010.03.006

[2] Antonio Alcaro and Jaak Panksepp. 2011. The SEEKING mind: Primal neuroaffective substrates for appetitive incentive states and their pathological dynamics in addictions and depression. *Neuroscience & Biobehavioral Reviews* 35, 9 (Oct. 2011), 1805–1820. doi:10.1016/j.neubiorev.2011.03.002

[3] A Anthony and L Schüler. 2022. Characterising the Basic Emotions in Long COVID. Supervised by M. Solms, D. Minné and A. Y. Turan.

[4] AP Association and others. 2013. Diagnostic and Statistical Manual of Mental Disorders. American Psychiatric Association. *Arlington, VA* (2013).

[5] R.M. Baños, C. Botella, M. Alcañiz, V. Liaño, B. Guerrero, and B. Rey. 2004. Immersion and Emotion: Their Impact on the Sense of Presence. *CyberPsychology & Behavior* 7, 6 (Dec. 2004), 734–741. doi:10.1089/cpb.2004.7.734

[6] Aaron T. Beck, Robert A. Steer, and Gregory K. Brown. 1996. *Manual for the Beck Depression Inventory-II.* Psychological Corporation, San Antonio, TX.

[7] Margaret M. Bradley and Peter J. Lang. 1994. Measuring emotion: The self-assessment manikin and the semantic differential. 25, 1 (1994), 49–59. doi:10.1016/0005-7916(94)90063-9

[8] Margaret M. Bradley and Peter J. Lang. 2007. *The International Affective Digitized Sounds (2nd Edition; IADS-2): Affective Ratings of Sounds and Instruction Manual.* Technical Report Technical Report B-3. University of Florida, Gainesville, FL.

[9] Alice Chirico, Francesco Ferrise, Lorenzo Cordella, and Andrea Gaggioli. 2018. Designing Awe in Virtual Reality: An Experimental Study. *Frontiers in Psychology* 8 (Jan. 2018), 2351. doi:10.3389/fpsyg.2017.02351

[10] Nicolò Dozio, Federica Marcolin, Giulia Wally Scurati, Francesca Nonis, Luca Ulrich, Enrico Vezzetti, and Francesco Ferrise. 2021. Development of an affective database made of interactive virtual environments. *Scientific Reports* 11, 1 (Dec. 2021), 24108. doi:10.1038/s41598-021-03380-y

[11] Nicoló Dozio, Federica Marcolin, Giulia Wally Scurati, Luca Ulrich, Francesca Nonis, Enrico Vezzetti, Gabriele Marsocci, Alba La Rosa, and Francesco Ferrise. 2022. A design methodology for affective Virtual Reality. *International Journal of Human-Computer Studies* 162 (June 2022), 102791. doi:10.1016/j.ijhcs.2022.102791

[12] Anna Felnhofer, Oswald D. Kothgassner, Mareike Schmidt, Anna-Katharina Heinzle, Leon Beutl, Helmut Hlavacs, and Ilse Kryspin-Exner. 2015. Is virtual reality emotionally arousing? Investigating five emotion inducing virtual park scenarios. *International Journal of Human-Computer Studies* 82 (Oct. 2015), 48–56. doi:10.1016/j.ijhcs.2015.05.004

[13] Miyo Hori, Kazuo Yamada, Junji Ohnishi, Shigeko Sakamoto, Hiroki Furuie, Kazuo Murakami, and Yukio Ichitani. 2014. Tickling during adolescence alters fear-related and cognitive behaviors in rats after prolonged isolation. *Physiology & Behavior* 131 (May 2014), 62–67. doi:10.1016/j.physbeh.2014.04.008

[14] Peter J. Lang, Margaret M. Bradley, and Bruce N. Cuthbert. 2008. *International Affective Picture System (IAPS): Affective Ratings of Pictures and Instruction Manual.* Technical Report Technical Report A-8. University of Florida, Gainesville, FL.

[15] Joomi Lee, Chen Crystal Chen, and Aryabrata Basu. 2025. From Novelty to Knowledge: A Longitudinal Investigation of the Novelty Effect on Learning Outcomes in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics* 31, 5 (May 2025), 3204–3212. doi:10.1109/TVCG.2025.3549897

[16] Yisi Liu and Olga Sourina. 2013. EEG Databases for Emotion Recognition. In *2013 International Conference on Cyberworlds* (Yokohama, Japan, 2013-10). IEEE, 302–309. doi:10.1109/CW.2013.52

[17] Christian Montag, Katharina Widenhorn-Müller, Jaak Panksepp, and Markus Kiefer. 2017. Individual Differences in Affective Neuroscience Personality Scale (ANPS) Primary Emotional Traits and Depressive Tendencies. *Comprehensive Psychiatry* 73 (2017), 136–142. doi:10.1016/j.comppsych.2016.11.007

[18] Kristine L. Nowak and Frank Biocca. 2003. The Effect of the Agency and Anthropomorphism on Users' Sense of Telepresence, Copresence, and Social Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments* 12, 5 (2003), 481–494. doi:10.1162/105474603322761289

[19] Jaak Panksepp. 2005. Affective consciousness: Core emotional feelings in animals and humans. *Consciousness and Cognition* 14, 1 (March 2005), 30–80. doi:10.1016/j.concog.2004.10.004

[20] Jaak Panksepp. 2010. Affective neuroscience of the emotional BrainMind: evolutionary perspectives and implications for understanding depression. *Dialogues in Clinical Neuroscience* 12, 4 (Dec. 2010), 533–545. doi:10.31887/DCNS.2010.12.4/jpanksepp

[21] Jaak Panksepp and Lucy Biven. 2012. *The archaeology of mind: neuroevolutionary origins of human emotions* (1st ed ed.). W. W Norton.

[22] Jaak Panksepp, Steve Siviy, and Larry Normansell. 1984. The psychobiology of play: Theoretical and methodological perspectives. *Neuroscience & Biobehavioral Reviews* 8, 4 (Dec. 1984), 465–492. doi:10.1016/0149-7634(84)90005-8

[23] Lisa Rebenitsch and Charles Owen. 2014. Individual variation in susceptibility to cybersickness. In *Proceedings of the 27th annual ACM symposium on User interface software and technology.* ACM, Honolulu Hawaii USA, 309–317. doi:10.1145/2642918.2647394

[24] Giuseppe Riva, Fabrizia Mantovani, Claret Samantha Capideville, Alessandra Preziosa, Francesca Morganti, Daniela Villani, Andrea Gaggioli, Cristina Botella, and Mariano Alcañiz. 2007. Affective Interactions Using Virtual Reality: The Link between Presence and Emotions. *CyberPsychology & Behavior* 10, 1 (Feb. 2007), 45–56. doi:10.1089/cpb.2006.9993

[25] James Russell and Albert Mehrabian. 1977. Evidence for a Three-Factor Theory of Emotions. *Journal of Research in Personality* 11 (Sept. 1977), 273–294. doi:10.1016/0092-6566(77)90037-X

[26] 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:10.1162/105474601300343603

[27] Gal Sheppes, Gaurav Suri, and James J. Gross. 2015. Emotion Regulation and Psychopathology. *Annual Review of Clinical Psychology* 11 (2015), 379–405. doi:10.1146/annurev-clinpsy-032814-112739

[28] Rukshani Somarathna, Tomasz Bednarz, and Gelareh Mohammadi. 2021. *Virtual Reality for Emotion Elicitation – A Review.* doi:10.48550/arXiv.2111.04461

[29] Rukshani Somarathna, Tomasz Bednarz, and Gelareh Mohammadi. 2023. Virtual Reality for Emotion Elicitation – A Review. *IEEE Transactions on Affective Computing* 14, 4 (Oct. 2023), 2626–2645. doi:10.1109/TAFFC.2022.3181053

[30] Lisa Stempfer, Sarah E. M. Stoll, Jonathan Fries, Reinhard Pekrun, and Thomas Goetz. 2025. A systematic review and multilevel meta-analysis of the relationship between boredom and arousal. *Communications Psychology* 3, 1 (Nov. 2025), 182. doi:10.1038/s44271-025-00358-8

[31] Tanh Quang Tran, Tobias Langlotz, Jacob Young, Thomas W. Schubert, and Holger Regenbrecht. 2024. Classifying Presence Scores: Insights and Analysis from Two Decades of the Igroup Presence Questionnaire (IPQ). *ACM Transactions on Computer-Human Interaction* 31, 5, Article 61 (2024), 26 pages. doi:10.1145/3689046

[32] David Watson and Lee Anna Clark. 1994. The PANAS-X: Manual for the Positive and Negative Affect Schedule - Expanded Form. Institution: University of Iowa. doi:10.17077/48vt-m4t2

A Summary Statistical Output

The summary statistics for all hypotheses is shown in Table 1 and descriptive results for relevant metrics in Table 2.

B Further Details on H-TEL-1

In test H-TEL-1, we aimed to determine if participant throws were made in response to NPC throws, indicating a temporal alignment and as such, a reciprocal throwing nature. We captured the number of times a participant made at least one throw in between two successive NPC throws, and classified such throw as an answered throw. We then calculated the proportion of NPC throw intervals that were responded to. However, there is a chance that participant throws just so happened to fall within these NPC intervals, and were not made in temporal alignment with the NPC. To account for this, we isolated the throwing pattern of each participant and created 1000 surrogate datasets where the throwing pattern was retained, but the timing was shifted by some constant. These surrogate datasets would break any NPC throw temporal alignment. We then calculated the average answered proportion of NPC throws across all surrogates, which represents the proportion of NPC throws that would be answered purely due to the participants independent throw rhythm. Using this surrogate average, we subtracted it from the observed answered proportion to determine the answered proportion that could not be attributed to the participant's independent throw rhythm. If participants were throwing independent of the NPC, this value would be roughly 0.

Table 1: Summary of all hypothesis tests. p-values marked † are Holm-Bonferroni adjusted within their test group; unmarked values are unadjusted. Effect sizes are matched rank-biserial correlation (r) for Wilcoxon signed-rank tests and Cohen's d (d_z for paired designs) for t-tests. P2 and P3 denote the post-tutorial and post-affective readings. Telemetry was recorded for 23 of the 25 participants.

Hypothesis	Comparison	Test	Statistic	p	Effect size	Outcome
H-SAM-1v	Valence at P3 vs. 6.33	Wilcoxon	$V = 325.0$	$< .001^{\dagger}$	$r = 1.00$	Supported
H-SAM-1a	Arousal at P3 vs. 6.33	Wilcoxon	$V = 206.0$	$.245^{\dagger}$	$r = .27$	Not supported
H-SAM-1d	Dominance at P3 vs. 6.33	Wilcoxon	$V = 230.0$	$.141^{\dagger}$	$r = .42$	Not supported
H-SAM-2v	Δ valence, P2 to P3	Wilcoxon	$V = 35.0$	$.134^{\dagger}$	$r = .56$	Not supported
H-SAM-2a	Δ arousal, P2 to P3	Wilcoxon	$V = 92.0$	$.217^{\dagger}$	$r = .35$	Not supported
H-SAM-2d	Δ dominance, P2 to P3	Wilcoxon	$V = 63.0$	$.517^{\dagger}$	$r = .20$	Not supported
H-PANASX-1	Δ Joviality vs. Δ other subscales, P2 to P3	Wilcoxon	$V = 236.0$	.050	$r = .45$	Supported
H-PANASX-2	Joviality vs. runner-up (Gen. Pos. Affect) at P3	Paired t	$t(24) = 2.15$	$.042^{\dagger}$	$d_z = 0.43$	Supported
H-ANSS-1	Δ PLAY vs. Δ other subscales, P2 to P3	Paired t	$t(24) = 3.26$	.003	$d_z = 0.65$	Supported
H-ANSS-2	PLAY vs. runner-up (SEEKING) at P3	Wilcoxon	$V = 84.5$	$.718^{\dagger}$	$r = .10$	Not supported
H-CP-1	Copresence at P3 vs. midpoint 3	Wilcoxon	$V = 306.5$	$< .001$	$r = .89$	Supported
H-SP-1	Social Presence at P3 vs. midpoint 4	One-sample t	$t(24) = 2.49$	.020	$d = 0.50$	Supported
H-TEL-1	Observed vs. surrogate answered proportion	One-sample t	$t(22) = 1.32$	.202	$d = 0.27$	Not supported
H-iPQ-1	iPQ at P3 vs. midpoint 0	One-sample t	$t(24) = 10.20$	$< .001$	$d = 2.04$	Supported
H-iPQ-2	Δ iPQ, P2 to P3	Paired t	$t(24) = 2.46$	.021	$d_z = 0.49$	Supported

Table 2: Descriptive statistics for all measures at the readings used by the stipulated hypotheses. P1–P3 denote the pre-tutorial, post-tutorial and post-affective readings. PANAS-X subscales are mean item scores (1–5); ANSS subscales are sums (5–25); the iPQ is scored −3 to +3. The reciprocity index is the observed minus surrogate answered proportion ($n = 23$).

Measure	Reading	*Mdn*	*M*	*SD*
SAM Valence	P1	7	7.06	1.23
	P2	8	8.10	0.87
	P3	8	8.32	0.75
SAM Arousal	P1	5	4.82	2.04
	P2	7	6.22	2.27
	P3	7	6.68	2.34
SAM Dominance	P1	6	6.08	1.57
	P2	7	6.70	1.49
	P3	7	6.88	1.56
Joviality	P3	4.38	4.18	0.78
General Positive Affect	P3	4.00	3.97	0.68
Attentiveness	P3	4.00	3.79	0.76
Surprise	P3	3.67	3.77	0.87
Serenity	P3	3.67	3.72	0.68
Self-Assurance	P3	3.50	3.53	0.92
Fatigue	P3	1.25	1.40	0.52
Shyness	P3	1.25	1.33	0.39
Fear	P3	1.00	1.13	0.22
General Negative Affect	P3	1.00	1.09	0.12
Guilt	P3	1.00	1.08	0.22
Hostility	P3	1.00	1.07	0.14
Sadness	P3	1.00	1.06	0.18
PLAY	P3	21	20.16	4.00
SEEKING	P3	21	20.08	3.90
CARE	P3	17	16.16	4.84
PANIC	P3	5	5.52	1.19
GRIEF	P3	5	5.20	0.71
RAGE	P3	5	5.16	0.47
FEAR (ANSS)	P3	5	5.12	0.60
iPQ overall	P2	1.21	1.14	0.77
	P3	1.50	1.41	0.69
Copresence	P3	3.82	3.75	0.60
Social Presence	P3	4.50	4.58	1.16
Reciprocity index	—	0.014	0.009	0.032