

Using Virtual Reality to Elicit the PLAY Emotion

Design and Evaluation of an Apple Interaction Environment

Zenande Mdludlu
University of Cape Town
Cape Town, South Africa
mdlzen005@myuct.ac.za

Abstract

Panksepp's Theory of Affective Neuroscience identifies PLAY as a distinct, sub-cortical primary-process emotion, but no prior work has targeted PLAY elicitation in Virtual Reality (VR). This paper adapts an existing affective virtual environment (VE) design framework to incorporate Pankseppian theory, decomposing PLAY into a high valence-arousal-dominance (VAD) profile together with a qualitative requirement for social, reciprocal interaction. The framework is instantiated in an Apple Interaction VE, in which a participant and a squirrel-like non-playable character (NPC) freely throw and retrieve apple-like objects with no imposed goal. Twenty-five participants completed the environment, with subjective measures taken at three points, before any VR exposure, after a neutral tutorial, and after the affective environment, allowing changes attributable to VR in general to be separated from changes attributable to the target environment specifically. Relative to the pre-VR baseline, valence, arousal, PANAS-X joviality, and self-reported PLAY (via the Affective Neuroscience State Scales, ANSS) all increased significantly; however, this rise was concentrated in the transition into VR generally rather than the transition into the Apple environment specifically, and self-reported dominance showed no reliable change across any of the three stages once the pre-VR baseline was included. Presence and co-presence were significantly above their scale reference points, whereas social presence was not, and arousal did not exceed a pre-defined high threshold. Behavioural data showed no decline in engagement over time, and the proportion of participant-initiated interactions, rather than raw interaction volume, correlated significantly with self-reported PLAY. These results offer partial, and more conservative than initially estimated, evidence that the adapted framework can support PLAY elicitation in VR, and point to participant initiative, rather than engagement volume or dominance ratings, as the more informative design and measurement target going forward.

Keywords

Virtual Reality, Emotion Elicitation, Affective Neuroscience, PLAY Emotion, Primary-Process Emotions, Multimodal Measurement, Presence

1 Introduction

Emotions shape how people think, remember, and act, and understanding them experimentally requires being able to reliably bring a target emotion about in a controlled setting, a process generally called emotion elicitation [8]. Two broad traditions dominate this work. Discrete models treat emotions as a small set of distinct

categories (e.g. joy, fear, anger); dimensional models instead place any emotional state on continuous scales, most commonly valence (how pleasant or unpleasant an experience feels), arousal (how energised or calm it feels), and dominance (how much control over the situation it feels like one has), together abbreviated VAD. Jaak Panksepp's Theory of Affective Neuroscience offers a third, biologically grounded account, proposing that mammals possess a small number of innate, sub-cortical emotional circuits. Among these is PLAY: a non-serious, highly pleasurable form of back-and-forth interaction with another agent that does not require higher (neo-cortical) reasoning to occur, and is therefore classed as a primary-process emotion, closer to a built-in reflex than a reasoned response [29–31].

Historically, emotions have been elicited using static images, film clips, or autobiographical recall, methods in which the participant remains a passive observer of content shown to them. This limits the ecological validity of the resulting state, i.e. how well it resembles the emotion as it actually occurs in the real world, and makes such methods poorly suited to an inherently interactive emotion like PLAY [8]. Active techniques, of which Virtual Reality (VR) is the most immersive available example, instead let the participant do something, and tend to produce more intense, embodied emotional responses [34, 38]. Two properties of VR are central to this effect. Agency is the degree to which the participant's own actions meaningfully change what happens in the environment, rather than the environment unfolding regardless of what they do. Presence is the subjective sense of "being there", of the virtual environment (VE), a computer-generated scene the participant experiences as if physically inside it, feeling real rather than merely watched. Because PLAY specifically requires ongoing, embodied engagement with another agent (in this context, a computer-controlled character, or non-playable character, NPC, that a participant can interact with), VR is a plausible medium for eliciting it; yet, to our knowledge, no existing VR study has specifically targeted PLAY as a distinct neurobiological system rather than a generic positive, high-arousal state.

Two further, compounding problems motivate this paper's approach. First, existing frameworks for designing affective VEs are built around dimensional or discrete emotion models and do not encode the reciprocal, role-reversing interaction that defines PLAY: a VE built only to hit a target VAD region cannot be assumed to elicit PLAY specifically, since other high-valence, high-arousal states, such as SEEKING (Panksepp's term for appetitive, exploratory drive), occupy much the same region of VAD space. Second, existing self-report and physiological instruments were not designed to isolate PLAY: the Self-Assessment Manikin (SAM), a widely used

pictorial VAD rating scale, and similar instruments cannot distinguish PLAY from other high-arousal positive states, and physiological signals such as heart rate indicate general arousal rather than any one specific emotional system. Consequently, even a well-designed PLAY-oriented VE cannot be convincingly validated using any single existing measure, motivating both a design framework that explicitly encodes PLAY's qualitative, social character, and an evaluation approach that triangulates across several imperfect measures rather than relying on any one of them.

This paper addresses that gap by adapting Dozio et al.'s affective VE design framework [12] to incorporate Pankseppian theory, and by applying it to the design of a single VE: an Apple Interaction environment in which a participant and a squirrel-like NPC engage in open-ended, reciprocal object play, with no score, timer visible to the participant, or externally imposed goal. The environment and framework are evaluated against the following questions:

- (RQ1):** To what extent does a VE specifically designed to target PLAY successfully elicit this primary-process system?
- (RQ2):** What design elements are required to afford PLAY behaviours, such as reciprocity, role reversal, and non-serious engagement, in VR?
- (RQ3):** Can triangulation of self-report, presence, and behavioural measures infer state-level PLAY activation?

The remainder of this paper reviews the theoretical and empirical background needed to situate these questions (Section 2), presents the adapted Pankseppian design framework and its implementation in the Apple Interaction VE (Section 3), describes the experimental design used to evaluate it (Section 4), and reports (Section 5) and discusses (Section 6) the resulting evidence for PLAY elicitation before concluding (Section 7).

2 Related Work

2.1 Foundations: Emotion Theories and Virtual Reality

Before turning to work specific to VR and PLAY, it is useful to fix a small amount of shared vocabulary, since the remainder of this paper draws on several adjacent but distinct traditions. As introduced above, discrete and dimensional models offer two established ways of describing an emotional state; Panksepp's affective neuroscience offers a third, arguing that a handful of primary-process systems, PLAY among them, are built into the mammalian brain below the level of conscious appraisal [27]. These three traditions are not mutually exclusive: a state can simultaneously be described by which primary-process system is active, where it falls on VAD, and what discrete label a person might give it, and this project uses all three descriptions at different points, dimensional VAD to specify design targets (Section 3), and both discrete labels and Pankseppian systems to interpret results (Sections 5 and 6).

Within VR specifically, presence and agency, introduced above, are the two properties most consistently linked to how strongly an emotion is felt [34]. A further property, immersion, refers to the objective, technical capacity of a VR system to isolate the senses (e.g. a head-mounted display blocking out the physical room), which is a precondition for, but conceptually distinct from, the participant's subjective sense of presence: a technically immersive headset does

not guarantee that any given participant will feel present in the specific scene it displays [34]. With this vocabulary established, the remaining three subsections review, in turn, why VR is an effective elicitation medium generally, how affective VEs have previously been designed, and how emotion has previously been measured in this context, each closing on the specific gap left open for a Pankseppian, PLAY-targeted VE.

2.2 VR as an Active Elicitation Medium

VR is consistently found to be a more effective medium for controlled emotion elicitation than 2D displays or 360-degree video, producing higher arousal, presence, and positive affect [24], and VR-elicited responses are comparable in intensity and temporal pattern to real-world experiences [7, 35]. This efficacy is grounded in a bidirectional relationship between presence and emotion: emotionally charged environments increase presence, and higher presence in turn amplifies the resulting arousal [14, 34]. This bidirectionality matters for design, since it implies that simply making an environment more immersive is not sufficient on its own; the content experienced once present must itself carry the intended emotional signal. Interactive VR further outperforms passive viewing, an effect that is itself mediated by presence [4, 47], and comparative work across delivery formats finds a consistent gradient in which fully interactive VR outperforms 360-degree video, which in turn outperforms conventional 2D video, on measures of presence, enjoyment, and emotional engagement [46]. Taken together, this establishes VR, and specifically interactive VR, as an appropriate platform for investigating an inherently active, primary-process emotion such as PLAY, rather than a merely visually rich one.

2.3 Designing Affective, Social Virtual Environments

Design choices in a VE directly shape the emotion it elicits. Dozio et al. [12] formalised a four-layer methodology for affective VE design, validated across ten VEs, in which semantic (thematic), sensory, and interactive elements are each selected to target a specific VAD region; consistent with this, other work identifies theme, colour, lighting, sound, and haptics (touch-based feedback) as central levers for shaping emotional response [14, 18, 19, 39]. However, this and comparable frameworks have only been validated against dimensional or discrete emotion targets: none incorporate reciprocal social interaction or role-reversal, the behavioural signatures of PLAY, and their "interactive" layer emphasises user control over the environment (e.g. moving objects, solving a puzzle) rather than social reciprocity with an agent that itself has its own initiative.

Separately, VR research on social and collaborative interaction shows that a cooperative, agent-based companion enhances social presence and positive affect [10, 33, 44], particularly when the companion is a stylised, non-human character rather than a realistic human avatar [11], whereas an evaluative, rather than collaborative, agent (one that appears to be judging the participant's performance) can reduce positive affect [41]. Active, agent-directed tasks also produce a wider spread of arousal responses than passive viewing [45]. Yet this literature, like the design frameworks above, is almost uniformly task- and goal-oriented, e.g. escaping a room or completing an objective together, which is incompatible with

PLAY’s autotelic character, meaning an activity pursued for its own sake rather than towards an external goal. Taken together, this body of work establishes VR and reciprocal agent interaction as promising ingredients for PLAY elicitation, but leaves open how they should be combined and evaluated for a Pankseppian, rather than dimensional or goal-directed, emotional target, the gap this paper addresses.

2.4 Measuring Emotion in VR

Self-report instruments dominate affective VR evaluation but are poorly suited to isolating state-level PLAY. The SAM [5] conflates PLAY with SEEKING within the same high-valence, high-arousal region of VAD space; the PANAS-X Joviality subscale, part of the broader Positive and Negative Affect Schedule [43], reflects positive engagement generally rather than PLAY specifically; and self-report more broadly is susceptible to demand characteristics (participants unconsciously answering in the way they believe is expected of them) and other cognitive biases [20, 38]. Presence and social-presence instruments, such as the iGroup Presence Questionnaire (IPQ) [36] and the co-presence and social-presence subscales of Nowak and Biocca [25], capture the sense of “being there” and of interacting with a social other, respectively, but are explanatory rather than direct measures of PLAY itself: high presence tells us the experience felt real, not that it felt playful. No existing self-report instrument was designed to isolate PLAY from other positive, high-arousal states, which motivates the multimodal, triangulated evaluation approach, combining several such imperfect measures with objective behavioural logging, adopted in this paper.

3 Design

An affective VE design framework is, at its core, a recipe for turning a target emotion into concrete choices about what a participant sees, hears, and can do inside a VE. This section first presents the theoretical framework used to translate Pankseppian PLAY into such choices, then describes how that framework was instantiated in the Apple Interaction environment evaluated in this paper.

3.1 A Pankseppian Design Framework for PLAY

Panksepp describes the emotion system as a three-tier hierarchy (Figure 1): a primary sub-cortical system, including PLAY and the other Pankseppian emotions, which operates automatically and does not require conscious thought; a secondary system of learned associations, in which previously neutral stimuli come to reliably trigger a primary emotion because they have repeatedly co-occurred with it; and a tertiary system of reflective, cortical appraisal, roughly corresponding to conscious, deliberate emotional reasoning. These three tiers are related by a bidirectional cascade, in which activation at any one level eventually triggers the others [30, p. 73].

This hierarchy determines which route is available for eliciting PLAY in VR. Direct, primary-level (route one) elicitation of PLAY requires physical social contact and is therefore not possible through a headset and controllers. Reflective, tertiary-level (route three) elicitation is also unsuitable, since PLAY consists of instinctual responses to a rapidly changing environment rather than deliberate reasoning about it [31]. This leaves route two: activating the

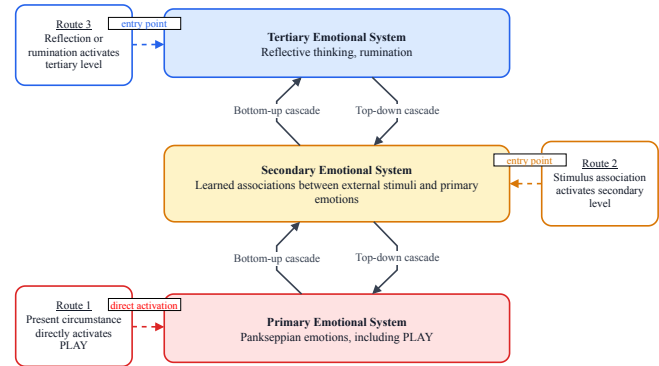


Figure 1: Panksepp’s three-tier emotional hierarchy (primary, secondary, tertiary systems), redrawn from Panksepp and Biven [30, p. 73].

secondary, learned-association layer with in-VE assets that participants already associate with playful behaviour (e.g. throwing objects back and forth with a responsive companion), from which activation cascades down to the primary PLAY system and, eventually, back up to the tertiary, dimensional level. This last step is what licenses the use of a dimensional design tool at all: because a triggered primary emotion is expected to eventually cascade up to the tertiary level, Dozio et al.’s VAD-based affective VE design framework [12] can be used to indirectly target PLAY, provided the selected design elements also carry the correct VAD profile.

PLAY is decomposed into a fast-paced, highly pleasurable, back-and-forth interaction in which each participant continuously reshapes and responds to the other’s actions. This corresponds to high valence, high arousal, and, since the participant continuously influences the shared environment, high dominance. Following prior VAD-based research, the continuous $[-1, 1]$ VAD scale is divided into three equal low/medium/high intervals [21], giving a target interval of $[0.33, 1]$ on each dimension. Because this VAD profile alone cannot distinguish PLAY from other positive, high-arousal states such as SEEKING, it is paired with a qualitative requirement, drawn directly from Panksepp’s account of PLAY, that the interaction be social and reciprocal in nature. This qualitative requirement is enforced through affective framing: the semantic and sensory content of the VE must not only carry the correct VAD ratings, but must also depict an ongoing, back-and-forth exchange with an NPC rather than a solitary high-arousal activity.

Because secondary-level activation additionally requires the in-VE stimulus to closely resemble its real-world conditioning stimulus, presence is treated as a design requirement rather than an incidental benefit. Presence is decomposed into immersion, agency, affective content quality, and uniformity of experience (consistency of visual and sensory style throughout, so that no single jarring element breaks the sense of being in one coherent place) [34], and each subcomponent is targeted at a specific stage of the design process: affective content quality through the affective framing described above, uniformity of experience through consistent semantic and sensory elements, agency through the interactive layer, and immersion through the fidelity of the VR hardware itself. Combining the

VAD and qualitative decomposition of PLAY with this presence-optimisation strategy yields the adapted, Pankseppian affective VE design framework used to design the environment described below (Figure 2), instantiated for PLAY specifically in Figure 3.

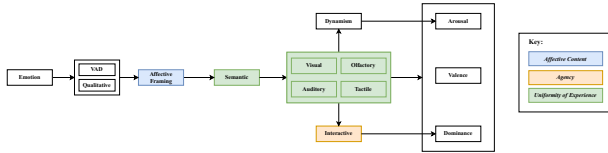


Figure 2: Adapted Pankseppian affective VE design framework, extending Dozio et al. [12] with a qualitative social-reciprocity component and explicit presence optimisation.

3.2 Environment Architecture and Implementation

The framework above was instantiated in the Apple Interaction VE (Figure 3), built in Unity using the XR Interaction Toolkit and OpenXR and structured as two scenes.

The tutorial scene, called VE_Tutorial, onboards the participant on the project’s XR rig, i.e. comfort-hop locomotion (a short, low-motion hop rather than continuous walking, chosen to reduce cybersickness, discussed below), and grabbing and throwing, driven by a step-based component called TutorialManager and a project-specific bridge component that binds these steps to the rig and loads the main scene on completion. Settings established here, including one of twelve hand skin tones chosen from a palette, persist into the main scene. The main scene, called Final, is a fixed 7.5-minute open-play session in a circular woodland meadow centred on a single large tree. A component called PlayEnvironmentBootstrap auto-wires the world, session timer, squirrel, apple manager, and logging systems. When the session timer elapses, the music fades, the scene fades to black, and locomotion is disabled.

Locomotion was deliberately comfort-oriented rather than free-roaming, using a short blink-hop in the participant’s look direction, with movement clamped to the meadow boundary, in order to reduce the risk of cybersickness (VR-induced nausea and disorientation) [32, 37] without resorting to a less embodied point-and-teleport arc. Grabbing and throwing are performed with the index trigger on objects built with the XR Interaction Toolkit’s grab interactable components. Apples are spawned in two ways: loose apples on the ground, capped at five and respawned at fixed meadow points as the pool drops below three, and apples growing on the tree itself, which detach into physics objects once grabbed. Both apple types use a custom shader that I created specifically for this project, together with star-particle sparkle effects (Figure 6), and use a single customisable material so that the apple’s colour can be changed; three colour presets were made for the study. The meadow’s grass similarly uses a custom shader I wrote, but its wind motion was removed to optimise performance (frame rate increased considerably after this change).

The squirrel NPC (Figure 5) is a NavMesh agent (an AI-controlled character that navigates the meadow’s walkable surface automatically) driven by a finite-state machine with a lightweight mood

model tracking “playfulness” and “tiredness”, which scale its movement speed and bias its behaviour over the session. Its sensors track the participant’s distance and idling together with the trajectory and hit volume of thrown apples, and its behaviour includes idling and wandering, offering an apple, chasing and catching thrown or loose apples with a catch style that varies by the apple’s height and side of approach, holding an apple before offering, evading in a keep-away pattern, dropping or throwing it, reacting to being thrown at (catching, dodging, a startle, a playful trip, or a taunt), watching the participant, and, uniquely to the squirrel rather than the participant, climbing the tree to perch and drop or toss apples from above. Participant-side tree-climbing was attempted in an early build via a chain of overlapping capsule colliders along the trunk, since the scaled tree mesh itself was unsuitable for climbing collision, but this remained unreliable and was non-functional for participants throughout data collection; the tree therefore served participants as a visual and social anchor rather than a climbable object, a limitation discussed further in Section 6. Expressive face textures, hand-drawn by me in Procreate, swap to reflect the squirrel’s mood, gaze direction, and blinking, and its animator drives corresponding walk/run, carry, catch/throw, climb, hit-reaction, and cheer animations; its vocalisations are pitch- and volume-jittered according to its current mood state. Audio more broadly consists of two-dimensional background music and ambience loops with apple impact sounds, all of which fade out softly at the end of the session. Throughout the session, logging managed by this same component writes discrete behavioural events, 10 Hz continuous pose and state data, and session metadata (proximity, following, idling, ignored offers, keep-away episodes, throws, and hand-off transfers) to CSV for later analysis; this logging is observation-only and does not alter gameplay, and Section 4.4 gives the precise operational definition of each logged indicator used in the analysis.

3.3 Sensory and Interaction Design Rationale

Each sensory and interaction choice above was treated as an affective-design hypothesis, intended to make the session feel safe, energetic, cheerful, and socially engaging rather than merely realistic, and each is grounded in prior VR and affective-computing findings rather than aesthetic preference alone.

The natural, open woodland setting follows evidence that natural virtual environments produce more positive cognitive and affective responses than urban ones [23] and that immersion in greenery, woodland, and semi-natural grassland is a recognised design element for positive virtual environments [39]. The open, circular, unobstructed meadow, together with daytime lighting, was likewise chosen for its association with openness, safety, and pleasantness, and to avoid the overly restrictive movement and threatening forms that the same guidelines advise against [39]. The custom grass shader was written specifically to give the environment a soft, non-abrasive visual texture, consistent with evidence linking soft and fuzzy textures to warmth, serenity, and joy [39]. The three apple colour presets were chosen to include warm, saturated hues, since bright colour is associated with joy in the same guidelines [39].

The background music loop was selected at a moderate-fast tempo, since faster tempo and major tonality are associated with

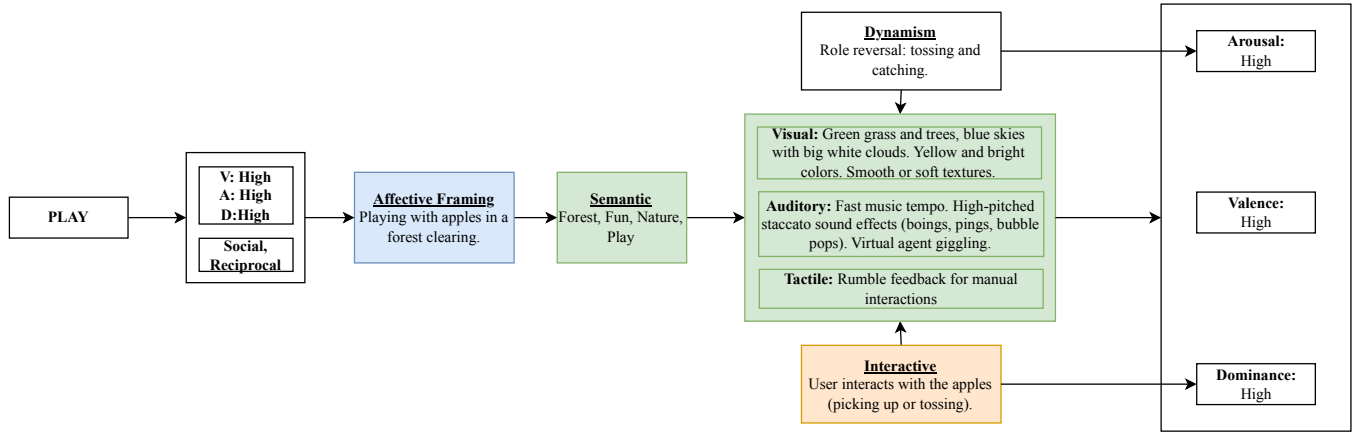


Figure 3: Apple Interaction VE design, following the adapted framework in Figure 2.



Figure 4: The Apple Interaction environment in session.



Figure 5: The squirrel NPC model in Blender.

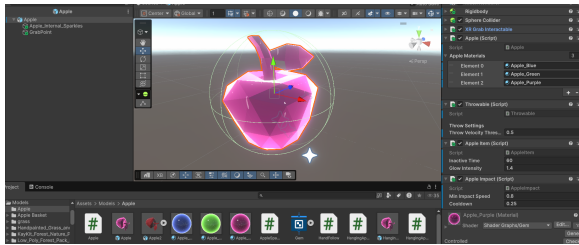


Figure 6: The apple prefab, with its custom shader, sparkle particle effect, and customisable material.

judgements of greater happiness [15]. This pairing of music with ambient outdoor sound follows a VR park study in which the presence of any sound increased presence relative to silence and music

was more effective than a nature soundscape at inducing an intended emotion [16], and closely mirrors the strongest existing precedent for this design, a VR joy-induction study that combined a cheerful tune with birdsong and chimes in a grassy park and tree grove [4]. Short, bright, high-register NPC sounds follow evidence that faster, higher-register sounds are perceptually associated with brighter, more energetic qualities in music [2].

The squirrel’s mood-linked vocal pitch and volume jitter, and its hand-drawn, swappable facial expressions, are intended as social rather than purely decorative cues: expressive behaviour, gaze, and vocal intonation are known to make virtual agents more believable and socially meaningful [42], and pitch and timbre specifically contribute to how happy or pleasurable a voice is judged to be [26]. This framing treats the squirrel’s apparent amusement and responsiveness as social reinforcement for the participant’s own play behaviour, consistent with virtual-pet research in which responsive, emotionally expressive animal behaviour, including bringing an object to the player, helped initiate and sustain engagement [17]. Tactile feedback on grab and catch events, which related work suggests can increase presence and a sense of emotional connection to a responsive virtual animal [17, 22], was considered but not implemented in the final build: informal observations I recorded during the experimental sessions described hand-controller haptics as distracting rather than engaging in this particular environment, and this was treated as a deliberate design decision rather than an oversight, though it may be a possible change for future iterations. As with other unverified observations from data-collection sessions reported later in this paper, this single design-time observation should be treated as suggestive rather than conclusive, and would benefit from a systematic follow-up (e.g. a controlled A/B comparison of haptic and non-haptic builds) before being generalised.

4 Experimental Design

This experiment evaluates whether the Apple Interaction VE, as designed in Section 3, elicits PLAY, and whether the resulting subjective, presence, and behavioural measures agree with one another. Subjective measures were collected at three points: before any VR exposure (pre-tutorial), after the neutral tutorial VE (post-tutorial),

and after the Apple Interaction VE itself (post-affective). This three-stage design, rather than a single pre/post comparison, was adopted specifically so that the pre-tutorial (initial) questionnaire could be used as a genuine, no-VR-exposure baseline in the statistical analysis, separating any general effect of simply entering VR (captured by the pre-tutorial to post-tutorial change) from any effect specific to the Apple environment (captured by the post-tutorial to post-affective change); Section 4.5 details how this is tested. Unlike in the project proposal, this evaluation does not include physiological (ECG or GSR) measures: these were removed from the final protocol, both to reduce session length and setup burden and because physiological arousal alone cannot distinguish PLAY from other high-arousal states, so their absence is compensated for by the additional ANSS and behavioural measures described below.

4.1 Ethics

This study received ethical clearance from the University of Cape Town Inter-Faculty Human Research Ethics Committee (IFHREC), IFHREC/02979/2026. Informed consent was obtained from all participants prior to any involvement; the consent process outlined the general VR-related risks (see below), the study procedure, and participants' right to withdraw at any time without consequence. The specific emotion under investigation, PLAY, was not disclosed to participants in order to avoid response bias; participants were told only that the study concerned their experience of a VR environment. This constitutes a mild form of deception by omission, since disclosing the true target emotion in advance would plausibly bias participants toward performing, rather than spontaneously exhibiting, playful behaviour, undermining the validity of the very measures used to detect it. This omission was judged minimal-risk, since no participant was misled about anything that could affect their safety, comfort, or willingness to take part, and it was fully addressed at debrief: participants were told the study's actual purpose, given the opportunity to ask questions, and offered the chance to withdraw their data after learning this, none of whom chose to do so. Key VR risks addressed in the consent process were cybersickness and post-exposure fatigue [37], whose likelihood is known to correlate with a prior history of motion sickness or migraines [32]. All data was stored securely in accordance with the Protection of Personal Information Act (POPIA) and the ACM Code of Ethics.

4.2 Participants

Participants exclusion criteria comprised existing psychiatric disorders, chronic stimulant use, a Beck Depression Inventory (BDI) score above 20 [3], and a history of motion sickness. Each criterion targets a specific, identifiable threat to the validity of a PLAY-elicitation claim rather than being a generic precaution. Existing psychiatric disorders and depression (indexed by the BDI) are excluded because both are independently associated with reduced expression of social play behaviour [29, 31]: including such participants would risk a false negative, in which the environment is capable of eliciting PLAY but a participant's baseline emotional state suppresses the behavioural and subjective signal regardless. Chronic stimulant use is excluded for a related reason, since psychostimulant medication is likewise linked to reduced playfulness [28], and its presence would introduce an uncontrolled source of

Table 1: Experimental procedure for the Apple Interaction VE.

Activity	Duration	Notes
Consent & pre-tutorial questionnaire	10 min	Baseline, before any VR exposure
Headset fitting & Tutorial VE	5 min	Baseline navigation and interaction
Post-tutorial questionnaires	5 min	SAM, PANAS-X, ANSS, IPQ
Apple Interaction VE	7 min	Behavioural events recorded
Post-affective questionnaires	10 min	SAM, PANAS-X, ANSS, IPQ, Co-Presence, Social Presence
Debrief	5 min	Deception disclosed; participant may leave when ready

between-participant variance in exactly the measures this study relies on. A history of motion sickness is excluded because it is a known predictor of cybersickness [32], which would both raise the ethical risk of harm to susceptible participants and confound self-reported arousal and valence with VR-induced nausea rather than PLAY-relevant affect. Taken together, these criteria are best understood as protecting the study's internal validity (by removing known sources of measurement confound) and the wellbeing of participants who would otherwise face disproportionate risk, rather than as excluding any group without a specific, stated methodological reason.

An a priori power analysis (medium effect size $d = 0.5$, $\alpha = .05$, power = .80, G*Power [9, 13]) indicated a target of 30 participants; this environment was completed by a final sample of $N = 25$ participants, short of the planned target. No participants were excluded post-hoc or withdrew after starting the study.

4.3 Procedure

Table 1 summarises the session structure. Participants first provided informed consent and completed the pre-tutorial questionnaire, then were fitted with the head-mounted display before completing the neutral tutorial VE, immediately followed by the post-tutorial questionnaire battery. Participants then completed the Apple Interaction VE, followed by the full post-affective questionnaire battery, before being debriefed. To minimise my own effect on participants' behaviour as the person running each session, sessions followed a standardised script [8].

4.4 Measures

Table 2 summarises the measures used and what each was intended to capture; each was selected to target one part of the PLAY decomposition in Section 3.1, or to serve an explanatory role, and the case for why a significant result on each measure matters is given alongside it below, since a measure is only worth including if its outcome would actually change what can be concluded about PLAY.

The SAM [5] directly measures valence, arousal, and dominance, providing the most direct test of PLAY’s VAD profile. Elevated valence matters because pleasure is definitional to PLAY: a non-pleasurable interaction cannot be PLAY regardless of what else is true of it. Elevated arousal matters because it is what distinguishes PLAY’s boisterous, fast-paced character from calmer positive states such as CARE; without it, a positive result would be equally consistent with quiet contentment. Elevated dominance matters because it operationalises the participant’s continuous influence over the interaction, PLAY as something one actively does together with an agent, not something pleasant that merely happens to a passive recipient. The PANAS-X Joviality subscale, part of the broader Positive and Negative Affect Schedule [43], was included as a validated, independent measure of joyful positive affect that does not share the SAM’s single-item, pictorial-scale method, so that convergence between the two reduces the chance that a valence result is an artefact of one instrument’s particular format; testing Joviality specifically against the next-highest subscale (surprise, in this sample) additionally rules out the possibility that the environment’s positive effect is really an orienting or novelty response rather than joy. The ANSS PLAY subscale is a state-level, Pankseppian self-report instrument developed by Solms and colleagues and is, to date, described only in an unpublished undergraduate research report [1]; it was not developed for this project, but is used here because it is the only instrument available that asks directly about Pankseppian primary-emotion activation rather than a generic dimensional or affect construct, making it, of all the measures used here, the only one that is explicitly Pankseppian rather than adapted from a general dimensional or affect framework; showing that PLAY specifically, rather than its closest and most easily confused neighbouring system, SEEKING, is the highest-rated ANSS subscale is the most direct test available that the target system, and not a nearby one, was activated. The co-presence and social-presence subscales developed by Nowak and Biocca [25] target PLAY’s qualitative, social-reciprocity requirement directly: co-presence asks whether the participant experienced the squirrel as a genuine social interactant rather than an inert object, which is the precondition for calling the exchange “social” at all, while social presence asks about the squirrel’s perceived attentiveness and engagement specifically, a distinct question, since the squirrel’s behaviour repertoire (offering, evading, chasing) could plausibly support a sense of another entity being present without that entity feeling genuinely attentive to the participant. The IPQ [36] was retained as an explanatory measure: since presence is linked to the intensity of emotion elicitation [14, 34], low presence offers a plausible alternative explanation for a null result on the primary measures, participants who did not feel present were arguably never given a fair chance to feel PLAY, rather than evidence that the framework itself failed.

Finally, behavioural indicators were logged automatically from gameplay telemetry to provide an objective, non-self-report window onto the reciprocal exchange itself, and, since these are reported in some technical detail in Section 5, their precise operational definitions are given here. Total interactions counts every discrete apple-related event (throw, catch, pickup, offer, or hand-off) recorded per participant across the session; interaction rate normalises this by session duration. Participant-initiated ratio and NPC-initiated ratio are the proportion of these events initiated

Table 2: Measures used to evaluate PLAY elicitation.

Measure	Type	Captures
Self-Assessment Manikin (SAM)	Subjective	Valence, Arousal, Dominance
PANAS-X (Joviality subscale)	Subjective	Joyful positive affect
ANSS (PLAY subscale)	Subjective	Pankseppian PLAY activation
iGroup Presence Questionnaire (IPQ)	Subjective	Presence (explanatory)
Co-Presence & Social Presence subscales	Subjective	Social reciprocity
Behavioural interaction log	Behavioural	Frequency, initiation, and chaining of apple exchanges

by the participant versus the squirrel respectively (and sum to 1), describing who is leading the exchange; this matters specifically because Panksepp’s account of PLAY emphasises mutual initiative and role-reversal, so if this ratio were unrelated to subjective PLAY it would suggest that initiative is not, in fact, what participants are keying into when they report feeling playful, which would weaken the framework’s specific emphasis on dominance and agency. A chain is a continuous, uninterrupted sequence of alternating throws and catches between participant and squirrel without a gap exceeding a fixed inactivity threshold; mean/max chain length and chain count jointly describe how sustained, rather than one-off, the reciprocal exchange was. Active-time ratio is the proportion of the session during which the participant was engaged in an interaction-relevant state rather than idle or wandering. Role-balance ratio measures how evenly interaction was distributed between initiating and receiving roles across the session (0.5 indicating a perfectly balanced back-and-forth), intended as a behavioural proxy for the framework’s role-reversal requirement. Time-in-proximity ratio is the proportion of session time spent within a fixed proximity threshold of the squirrel; this variable had an unexpectedly low mean (0.001) and near-zero median in the logged data, most likely reflecting an overly tight proximity threshold in the current logging implementation rather than participants genuinely almost never being near the squirrel, and its correlational results should therefore be read with this specific data-quality caveat in mind. Time-to-first-interaction is the latency, in minutes, to the participant’s first apple-related interaction.

4.5 Statistical Analysis

All subjective measures were reverse-coded where required and averaged according to their respective Likert scales prior to analysis. For SAM, PANAS-X, and ANSS, the inclusion of a genuine pre-VR (pre-tutorial) baseline allows a three-stage design: pre-tutorial, post-tutorial, and post-affective. For each subscale, an omnibus test, a single overall test for whether the three stages differ from one another at all, before asking which specific pair of stages differs, was run first, a repeated-measures ANOVA where the distribution of

stage-wise difference scores did not depart from normality (Shapiro-Wilk test), and its non-parametric equivalent, the Friedman test, otherwise. Only where this omnibus test was significant were the three pairwise stage comparisons (pre-tutorial vs. post-tutorial; pre-tutorial vs. post-affective; post-tutorial vs. post-affective) then tested, using paired *t*-tests or Wilcoxon signed-rank tests as appropriate, with a Holm-Bonferroni correction applied across the three comparisons within each subscale; this two-step procedure was used specifically to avoid inflating the false-positive rate that would result from testing all three pairwise comparisons regardless of whether any overall change had occurred. This design directly serves the purpose of including the initial (pre-tutorial) questionnaire in the analysis: the pre-tutorial vs. post-tutorial comparison isolates any effect of entering VR at all, independent of the Apple environment's specific content, while the post-tutorial vs. post-affective comparison isolates any effect attributable to the Apple environment specifically, over and above simply being in a headset.

Separately, and unaffected by this three-stage design, three single-timepoint comparisons from the post-affective assessment were retained from the original analysis plan: SAM valence, arousal, and dominance were each tested against a pre-defined "high" threshold (the upper third of the scale, consistent with prior dimensional-emotion research [21, 40]); PANAS-X Joviality was tested against its next-highest subscale; and ANSS PLAY was tested against its next-highest subscale. IPQ was tested for whether post-affective scores exceeded zero (see Section 5.3 for why zero is the relevant reference point for this instrument), and co-presence and social presence were each tested against their scale midpoints (see Section 5.3). Finally, for the relationship between behavioural engagement and subjective outcomes, Pearson or Spearman correlations (selected by the same normality criterion) were computed between each behavioural indicator and each subjective outcome, and a Friedman test was used to test for change in interaction frequency across the four two-minute bins of the Apple Interaction VE, addressing whether reciprocal engagement was sustained rather than front-loaded.

An alpha level of .05 was used throughout, with the Holm-Bonferroni correction described above applied within each family of related comparisons. Successful PLAY elicitation was defined as significance, in the predicted direction, on all primary measures across both the threshold and three-stage tests; partial support was defined as significance on a majority of primary measures together with converging evidence from at least one secondary or behavioural measure; and results were classified as inconclusive if post-affective presence fell significantly below baseline, since low presence offers an alternative explanation for any null result on the primary measures.

5 Results

Results are reported without interpretation and are organised around the three levels of evidence introduced in Section 4: subjective PLAY indicators, presence and social presence, and behavioural engagement, before a short summary against the elicitation criteria defined above. Interpretation of these results, including reconciling the more conservative three-stage findings below with the single-timepoint threshold tests, is deferred to Section 6.

5.1 Participant Characteristics

Twenty-five participants ($N = 25$) provided complete pre-tutorial, post-tutorial, and post-affective data for the Apple Interaction VE.

5.2 Subjective PLAY Indicators

Single-timepoint threshold tests. Post-affective SAM valence ($Mdn = 8$) and dominance ($Mdn = 7$) were both significantly above the high threshold (valence: Wilcoxon $V = 318, p < .001$; dominance: $V = 237, p = .023$), while arousal ($Mdn = 7$) was not ($V = 176, p = .363$). PANAS-X Joviality was the highest-rated subscale post-exposure ($M = 4.02$), ahead of surprise ($M = 3.72$), but this difference was not significant ($t(24) = 1.46, p = .079$). ANSS PLAY was the highest-rated ANSS subscale post-exposure ($M = 19.88$), significantly ahead of its closest competitor, SEEKING ($M = 18.80; t(24) = 2.25, p = .017$).

Three-stage analysis. Table 3 summarises the omnibus and, where applicable, pairwise three-stage results for the primary SAM and PANAS-X measures; the full three-stage results for every remaining PANAS-X and ANSS subscale, together with the full behavioural-subjective correlation matrix, are reported in Appendix C. SAM valence changed significantly across the three stages ($\chi^2(2) = 14.03, p < .001$; pre-tutorial $M = 7.16$, post-tutorial $M = 7.52$, post-affective $M = 8.08$); of the three pairwise comparisons, only pre-tutorial vs. post-affective survived correction ($p_{holm} = .017$), with neither pre-tutorial vs. post-tutorial ($p_{holm} = .306$) nor post-tutorial vs. post-affective ($p_{holm} = .160$) reaching significance. SAM arousal likewise changed significantly overall ($\chi^2(2) = 13.96, p < .001$; pre-tutorial $M = 4.42$, post-tutorial $M = 5.54$, post-affective $M = 6.18$), but here both the pre-tutorial vs. post-tutorial ($p_{holm} = .029$) and pre-tutorial vs. post-affective ($p_{holm} = .002$) comparisons were significant, while post-tutorial vs. post-affective was not ($p_{holm} = .116$). SAM dominance showed no significant omnibus change across the three stages ($\chi^2(2) = 4.59, p = .101$; pre-tutorial $M = 6.30$, post-tutorial $M = 6.44$, post-affective $M = 6.98$), and, following the two-step procedure in Section 4.5, no pairwise follow-up tests were run for this subscale.

PANAS-X Joviality changed significantly across the three stages ($\chi^2(2) = 20.18, p < .001$; pre-tutorial $M = 3.33$, post-tutorial $M = 3.78$, post-affective $M = 4.02$), with both pre-tutorial vs. post-tutorial ($p_{holm} = .003$) and pre-tutorial vs. post-affective ($p_{holm} < .001$) significant, but post-tutorial vs. post-affective not ($p_{holm} = .073$). PANAS-X surprise also changed significantly overall ($\chi^2(2) = 32.97, p < .001$; pre-tutorial $M = 2.25$, post-tutorial $M = 3.29$, post-affective $M = 3.72$), and here all three pairwise comparisons were significant, including post-tutorial vs. post-affective ($p_{holm} = .021$), the only comparison in Table 3 for which the Apple environment produced a further, statistically reliable increase beyond the neutral tutorial. ANSS PLAY changed significantly across the three stages by a repeated-measures ANOVA ($F(2, 48) = 9.43, p < .001$; pre-tutorial $M = 17.28$, post-tutorial $M = 19.16$, post-affective $M = 19.88$); no corrected pairwise contrasts were available for this subscale (see Appendix A). Figures 7a, 7b, and 8a in Appendix C show these SAM, PANAS-X, and ANSS distributions graphically.

Table 3: Three-stage change in primary subjective measures. p_{holm} given only where the omnibus test was significant.

Measure	Omnibus	Pre→Post-tut	Pre→Post-aff
SAM valence	$p < .001$	ns	$p_{holm} = .017$
SAM arousal	$p < .001$	$p_{holm} = .029$	$p_{holm} = .002$
SAM dominance	ns ($p = .101$)	—	—
PANAS-X Joviality	$p < .001$	$p_{holm} = .003$	$p_{holm} < .001$
PANAS-X Surprise	$p < .001$	$p_{holm} < .001$	$p_{holm} < .001$
ANSS PLAY	$p < .001$	(no corrected contrasts available)	

5.3 Presence and Social Presence

Presence and social presence were assessed differently across stages for a specific methodological reason: the IPQ can meaningfully be asked about any VE, since it concerns only the participant's own sense of being in a place, whereas the co-presence and social-presence subscales ask about interacting with another agent, and the neutral tutorial contains no NPC. Co-presence and social presence could therefore only be assessed once, post-affectively; this is a genuine limitation, addressed further in Section 6. IPQ, which does not require an agent, was assessed at both post-tutorial and post-affective stages.

The IPQ uses a bipolar response format centred on zero, where zero represents no sense of presence at all (a fully neutral, “not there” response) and increasingly positive scores indicate a stronger sense of being present in the VE [36]; testing whether post-affective scores exceed zero therefore tests whether participants felt any presence whatsoever, rather than a specifically high level of it, and is the appropriate first check before presence can be invoked as an explanatory factor at all. Post-affective IPQ scores were significantly above zero ($t(24) = 4.31, p < .001, M = 0.72, SD = 0.84$), but did not differ significantly from post-tutorial IPQ ($t(24) = 1.27, p = .108$, mean difference = 0.25), indicating that presence was established already during the neutral tutorial and did not increase further in the Apple environment specifically.

The co-presence and social-presence subscales instead use unipolar Likert scales with a defined neutral midpoint, a 5-point scale for co-presence (midpoint 3) and a 7-point scale for social presence (midpoint 4), at which a respondent neither agrees nor disagrees that the interaction felt social [25]; testing against this midpoint asks whether participants' average response leaned towards affirming, rather than denying or remaining neutral about, the social quality of the exchange with the squirrel, which is the more informative comparison for these particular instruments than a comparison against zero would be. Co-presence was significantly above its midpoint ($t(24) = 4.56, p < .001, M = 3.64, SD = 0.71$), whereas social presence did not differ significantly from its midpoint ($t(24) = -0.51, p = .691, M = 3.87, SD = 1.32$): participants tended to agree that the squirrel felt like a genuine social presence in the space with them, but were, on average, ambivalent about whether it felt specifically attentive or engaged with them.

5.4 Behavioural Engagement

Participants engaged in a mean of 193.4 apple-related interactions per session ($SD = 65.3$), of which a mean of 62.4% ($SD = 14.3\%$) were

participant-initiated (see Section 4.4 for the operational definition of each indicator below). A Friedman test found no significant change in interaction frequency across the four two-minute bins of the session ($\chi^2(3) = 2.65, p = .449$; bin means: 39.2, 47.3, 41.2, and 31.4 interactions), indicating that engagement was sustained rather than declining over the session. Of the six behavioural indicators tested against self-reported ANSS PLAY, only the participant-initiated ratio correlated significantly (Spearman $\rho = .52, p = .008$); total interactions, mean chain length, active-time ratio, role-balance ratio, and time-in-proximity ratio did not (all $p > .17$); the time-in-proximity result in particular should be read alongside the data-quality caveat for that variable noted in Section 4.4. This correlation was robust to the exclusion of any single participant (leave-one-out ρ range: .457–.615, p range: .0014–.025) and to removal of conventional outliers ($\rho = .578, p = .006$ with 4 IQR-flagged points excluded). The full 6×6 matrix of correlations between every behavioural indicator and every subjective outcome, including one further nominally significant but likely chance cell, is reported in the Appendix. Figures 8b, 9a, and 9b in Appendix A show these behavioural results, the temporal engagement pattern, and the outlier-robustness check graphically.

6 Discussion

Taken together, the results provide partial, and in one respect more conservative than the original single-baseline analysis suggested, support for the claim that the adapted, Pankseppian framework can guide the design of a VE that elicits PLAY. Valence and ANSS PLAY, the measure most directly tied to Panksepp's construct and the only explicitly Pankseppian instrument used here, both rose reliably from a genuine pre-VR baseline, and PLAY was rated above every competing primary-process subscale, including its closest neighbour, SEEKING. However, the three-stage design reveals a pattern a simpler two-point comparison would have missed: for SAM dominance, and arguably for valence and Joviality once the tutorial-to-affective step is examined alone, most of the measurable change occurred on entering VR at all, not on entering the Apple environment specifically. Dominance showed no reliable change once a genuine pre-VR timepoint was included, in contrast to the earlier, less conservative post-tutorial-baseline comparison; this is a meaningful revision, not a technicality, since dominance is one of the three VAD pillars the framework is built on. The only measure showing a further, reliable increase attributable to the Apple environment specifically was PANAS-X surprise, not Joviality, arousal, or dominance, suggesting the environment's most distinctive marginal contribution was an orienting response rather than PLAY itself. This, together with social presence failing to exceed its midpoint, complicates rather than confirms a clean PLAY-elicitation claim: a positive VR experience is not automatically the specific, socially reciprocal, high-arousal state the framework targets.

The behavioural evidence sharpens this cautious picture. Raw interaction frequency, chain length, and role balance were all unrelated to self-reported PLAY, but the proportion of interactions the participant initiated was significantly associated with it, a result that held up under both leave-one-out and outlier-removal checks (Appendix A) and is therefore not an artefact of one or two participants. This speaks directly to the design-element question raised in

the introduction: it is initiative within the exchange, not its volume, that matters for PLAY, a pattern invisible to a design process optimised only for engagement, and one at least suggestively consistent with the framework's emphasis on dominance, even though self-reported dominance itself did not survive the stricter three-stage test; behavioural initiative and self-reported dominance may simply tap the same construct with different reliability, worth testing directly with a larger sample. Session observations offer weaker, explicitly ad hoc supporting evidence, unsystematic notes rather than a structured protocol, so read as candidate explanations rather than findings: participants across multiple sessions tried to grab apples from the squirrel's hand, climb the tree, pet the squirrel unprompted, or kick the apple or squirrel. Read cautiously alongside the initiative result, these are at least consistent with a pull toward richer, more physical reciprocal interaction than the throw-and-catch loop could accommodate. The absence of any engagement decline across the four time bins further suggests the reciprocal, non-goal-directed structure sustained interest rather than depending on novelty, though this alone does not prove sustained playful engagement rather than habituated activity.

These findings carry several limitations, methodological and implementation-specific; the observational ones again carry the ad hoc caveat above. Physiological measures were dropped, both practically and because arousal alone cannot distinguish PLAY from other affective states, leaving the evidence entirely self-report and behavioural; future work could reintroduce physiological or neuroimaging measures to test whether they add information beyond ANSS and behavioural indicators, addressing PLAY's sub-cortical origin more directly. A structural limitation is that co-presence and social presence could only be measured once, post-Apple environment, since the neutral tutorial has no social agent for baselining; this leaves no way to isolate whether observed co-presence is specific to the squirrel or would occur with any first-encountered responsive agent. A future iteration could resolve this with a brief, non-playful scripted agent encounter during the tutorial (e.g., a stationary squirrel that merely acknowledges the participant) to provide a genuine non-PLAY social baseline. The sample ($N=25$) fell short of the target of 30, reducing power, and this paper covers one of three independently designed VEs in the broader project; cross-environment comparison, part of the original design, would clarify which specific design elements drove these results.

Several implementation limitations, again ad hoc and offered as candidate rather than confirmed explanations, plausibly capped achievable reciprocity. The squirrel sometimes threw from unrealistically close range, could not reliably hold or recover a second apple, and did not dodge throws, occasionally collapsing the exchange into one-sided fetch; a proximity-based grab mechanic, more consistent tree-climbing, a reaction to being hit while holding an apple, and support for multiple squirrels were all noted as targets for future iterations. Participant-side tree-climbing, attempted via capsule colliders along the trunk, remained unreliable and non-functional throughout data collection; since participants repeatedly asked about and attempted it, its absence likely constrained achievable dominance and exchange variety independently of the framework itself. Locomotion was not fully intuitive for everyone: some participants used a sliding motion instead of the intended blink-hop, some

struggled to combine movement and grabbing, and some had forgotten tutorial controls by the post-affective session; this, together with a boundary-rendering bug, suggests the tutorial baseline may not have been fully equivalent across participants. Future studies could replace the separate tutorial with a brief, experimenter-led live demonstration to reduce this risk. Finally, some participants were uncertain whether they were "supposed" to do something, and one was startled by the squirrel's appearance, suggesting the open-ended design may need a lighter-touch orientation rather than an explicit goal, and that individual differences in response to the NPC's design are worth monitoring.

For the framework itself, these results suggest the explicit qualitative reciprocity requirement was useful and identifiable in the behavioural data even where its self-report counterpart (dominance) did not survive stricter testing, and that participant-initiated interaction may be a more diagnostic PLAY signature than aggregate engagement or self-reported dominance. This is broadly consistent with my own methodological impression, again informal rather than tested, that behavioural measures may sit closer to Panksepp's construct than VAD-based self-report, given that instruments like the SAM were built for dimensional rather than Pankseppian models, whereas the explicitly Pankseppian ANSS showed the clearest theoretically direct evidence of PLAY-specific change. Future work could formalise the initiative-over-volume distinction, treating the participant-initiated ratio as a first-class design target alongside VAD, and developing a validated behavioural coding scheme for reciprocity and role-reversal that does not require manual session review. A more graceful ending, a gradual day-to-evening lighting transition in which the squirrel visibly settles down, rather than an abrupt fade to black, was also noted as a candidate refinement to close the exchange on an affectively consistent note.

7 Conclusion

This paper adapted an existing affective VE design framework to incorporate Panksepp's Theory of Affective Neuroscience and applied it to the design of an Apple Interaction VE intended to elicit the PLAY emotion. Using a three-stage design that included a genuine pre-VR baseline, valence, PANAS-X Joviality, and the explicitly Pankseppian ANSS PLAY subscale all rose reliably from that baseline, and PLAY was rated above every competing primary-process subscale; arousal did not exceed a pre-defined high threshold, and self-reported dominance showed no reliable change once the pre-VR baseline was included, a more conservative finding than an earlier two-point comparison suggested. Presence and co-presence were significantly above their reference points, while social presence was not, and behavioural data showed sustained engagement over the session together with a significant, outlier-robust link between the proportion of participant-initiated interaction, rather than total interaction volume, and self-reported PLAY. These results offer partial, convergent, but appropriately qualified evidence that a Pankseppian, reciprocity-aware adaptation of an existing affective VE design framework can support PLAY elicitation in VR, and point to participant initiative, together with explicitly Pankseppian self-report instruments such as the ANSS, as more informative targets for future design and measurement work in this space than aggregate engagement or dimensional dominance ratings alone.

AI Usage Statement

Generative Artificial Intelligence tools were used for proofreading and language editing of this manuscript (ChatGPT). All experimental design, analysis, and interpretation are the author's own.

References

- [1] Aaniyah Anthony and Luca Schüler. 2022. Characterising the Basic Emotions in Long COVID. (2022). https://humanities.uct.ac.za/sites/default/files/content_migration/humanities_uct_ac_za/1233/files/Aaniyah%2520Anthony%2520and%2520Luca%2520Schuler.pdf Unpublished Honours research report, Department of Psychology, University of Cape Town. Supervisors: Mark Solms, Donné Minné, Altay Yüce Turan.
- [2] Laura-Lee Balkwill and William Forde Thompson. 1999. A Cross-Cultural Investigation of the Perception of Emotion in Music: Psychophysical and Cultural Cues. *Music Perception* 17, 1 (Oct. 1999), 43–64. doi:10.2307/40285811
- [3] Aaron T. Beck, R. A. Steer, and G. Brown. 2011. Beck Depression Inventory–II. doi:10.1037/t00742-000 Institution: American Psychological Association.
- [4] Maor Bluman, Keren L.G. Snider, Guy Baratz, Amit Cohen, Daphna Canetti, and Béatrice S. Hasler. 2023. Virtual Reality-Based Joy Induction: The Role of Interactivity and Prior Mood. *Cyberpsychology, Behavior, and Social Networking* 26, 4 (April 2023), 229–237. doi:10.1089/cyber.2022.0250
- [5] 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
- [6] Margaret M. Bradley and Peter J. Lang. 2017. International Affective Picture System. *Encyclopedia of Personality and Individual Differences* (2017). doi:10.1007/978-3-319-28099-8_42-1
- [7] Alice Chirico and Andrea Gaggioli. 2019. When Virtual Feels Real: Comparing Emotional Responses and Presence in Virtual and Natural Environments. *Cyberpsychology, Behavior, and Social Networking* 22, 3 (March 2019), 220–226. doi:10.1089/cyber.2018.0393
- [8] James Coan and John Allen. 2007. *Handbook of Emotion Elicitation and Assessment*.
- [9] Jacob Cohen. 2013. *Statistical Power Analysis for the Behavioral Sciences* (0 ed.). Routledge. doi:10.4324/9780203771587
- [10] Julie Collange and Jérôme Guegan. 2020. Using virtual reality to induce gratitude through virtual social interaction. *Computers in Human Behavior* 113 (Dec. 2020), 106473. doi:10.1016/j.chb.2020.106473
- [11] Amir Dirin and Teemu Laine. 2023. The Influence of Virtual Character Design on Emotional Engagement in Immersive Virtual Reality: The Case of Feelings of Being. *Electronics* 12 (May 2023), 2321. doi:10.3390/electronics12102321
- [12] 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
- [13] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. 2007. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. 39, 2 (2007), 175–191. doi:10.3758/BF03193146
- [14] Anna Felnhöfer, Oswald D. Kothgassner, Mareike Schmidt, Anna-Katharina Heinze, 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
- [15] Patrick G. Hunter, E. Glenn Schellenberg, and Ulrich Schimmack. 2010. Feelings and Perceptions of Happiness and Sadness Induced by Music: Similarities, Differences, and Mixed Emotions. *Psychology of Aesthetics, Creativity, and the Arts* 4, 1 (2010), 47–56. doi:10.1037/a0016873
- [16] Angelika C. Kern, Wolfgang Ellermeier, and Lina Jost. 2020. The Influence of Mood Induction by Music or a Soundscape on Presence and Emotions in a Virtual Reality Park Scenario. In *Proceedings of the 15th International Audio Mostly Conference*. ACM, 233–236. doi:10.1145/3411109.3411129
- [17] Chaolan Lin, Travis Faas, and Erin Brady. 2017. Exploring Affection-Oriented Virtual Pet Game Design Strategies: Attachment, Motivations and Expectations of Users of Pet Games. In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. 362–369. doi:10.1109/ACII.2017.8273625
- [18] Dongbiao Luo. 2024. Design and Implementation of Virtual Reality Emotion Induction System. In *Proceedings of the 2024 International Conference on Intelligent Perception and Pattern Recognition*. ACM, Qingdao China, 1–5. doi:10.1145/3700035.3700036
- [19] Mariana Magalhães, Miguel Melo, António Fernando Coelho, and Maximino Bessa. 2024. Affective Landscapes: Navigating the Emotional Impact of Multisensory Stimuli in Virtual Reality. *IEEE Access* 12 (2024), 169955–169976. doi:10.1109/ACCESS.2024.3499858
- [20] Silvia Marocco, Valeria Vitale, Elena Grossi, Veronica Giffi, and Morena Santoriello. 2025. The impact of natural virtual environments on emotion elicitation: a state-of-the-art review. *Frontiers in Virtual Reality* 6 (Aug. 2025), 1638419. doi:10.3389/frvir.2025.1638419
- [21] J. Marín-Morales, J. Higuera-Trujillo, A. Greco, J. Guixeres, C. Llinares, E. Scilingo, M. Alcañiz, and G. Valenza. 2018. Affective computing in virtual reality: emotion recognition from brain and heartbeat dynamics using wearable sensors. *Scientific Reports* 8 (Sept. 2018). doi:10.1038/s41598-018-32063-4
- [22] Hye Sung Moon, Grady Orr, and Myounghoon Jeon. 2023. Hand Tracking with Vibrotactile Feedback Enhanced Presence, Engagement, Usability, and Performance in a Virtual Reality Rhythm Game. *International Journal of Human-Computer Interaction* 39, 14 (2023), 2840–2851. doi:10.1080/10447318.2022.2087000
- [23] Fariba Mostajeran, Jessica Krzikawski, Frank Steinicke, and Simone Kühn. 2021. Effects of Exposure to Immersive Videos and Photo Slideshows of Forest and Urban Environments. *Scientific Reports* 11, 1 (2021), 3994. doi:10.1038/s41598-021-83277-y
- [24] Yunfang Niu, Danli Wang, Ziwei Wang, Fan Sun, Kang Yue, and Nan Zheng. 2019. User Experience Evaluation in Virtual Reality based on Subjective Feelings and Physiological Signals. *Journal of Imaging Science and Technology* 63, 6 (Nov. 2019), 060413–1–060413–11. doi:10.2352/J.ImagingSci.Technol.2019.63.6.060413
- [25] 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. 12, 5 (2003), 481–494. doi:10.1162/105474603322761289
- [26] Christine Nussbaum, Annett Schirmer, and Stefan R. Schweinberger. 2022. Contributions of fundamental frequency and timbre to vocal emotion perception and their electrophysiological correlates. *Social Cognitive and Affective Neuroscience* 17, 12 (2022), 1145–1154. doi:10.1093/scan/nsac033
- [27] 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
- [28] Jaak Panksepp. 2007. Can PLAY diminish ADHD and facilitate the construction of the social brain? *Journal of the Canadian Academy of Child and Adolescent Psychiatry* 16, 2 (May 2007), 57–66.
- [29] 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
- [30] Jaak Panksepp and Lucy Biven. 2012. *The Archaeology of Mind: Neuroevolutionary Origins of Human Emotions* (1st ed.). W. W. Norton & Company.
- [31] Jaak Panksepp, Steve Sivi, 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
- [32] 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
- [33] Simon Riches, Soha Elghany, Philippa Garety, Mar Rus-Calafell, and Lucia Valmaggia. 2019. Factors Affecting Sense of Presence in a Virtual Reality Social Environment: A Qualitative Study. *Cyberpsychology, Behavior, and Social Networking* 22, 4 (April 2019), 288–292. doi:10.1089/cyber.2018.0128
- [34] 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
- [35] Radiah Rivu, Ruoyu Jiang, Ville Mäkelä, Mariam Hassib, and Florian Alt. 2021. Emotion Elicitation Techniques in Virtual Reality. In *Human-Computer Interaction – INTERACT 2021*, Carmelo Ardito, Rosa Lanzilotti, Alessio Malizia, Helen Petrie, Antonio Piccinno, Giuseppe Desolda, and Kori Inkpen (Eds.). Vol. 12932. Springer International Publishing, Cham, 93–114. doi:10.1007/978-3-030-85623-6_8 Series Title: Lecture Notes in Computer Science.
- [36] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. 2001. The Experience of Presence: Factor Analytic Insights. 10, 3 (2001), 266–281. doi:10.1162/105474601300343603
- [37] Beata Sokolowska. 2023. Impact of Virtual Reality Cognitive and Motor Exercises on Brain Health. *International Journal of Environmental Research and Public Health* 20, 5 (Feb. 2023), 4150. doi:10.3390/ijerph20054150
- [38] 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
- [39] Sophia C. Steinhäusser, Sebastian Oberdörfer, Sebastian Von Mammen, Marc Erich Latoschik, and Birgit Lugin. 2022. Joyful Adventures and Frightening Places-Designing Emotion-Inducing Virtual Environments. *Frontiers in Virtual Reality* 3 (July 2022), 919163. doi:10.3389/frvir.2022.919163
- [40] N. Suhaimi, J. Mountstephens, and J. Teo. 2022. A Dataset for Emotion Recognition Using Virtual Reality and EEG (DER-VREEG): Emotional State Classification Using Low-Cost Wearable VR-EEG Headsets. *Big Data and Cognitive Computing* 6 (Jan. 2022), 16. doi:10.3390/bdcc6010016

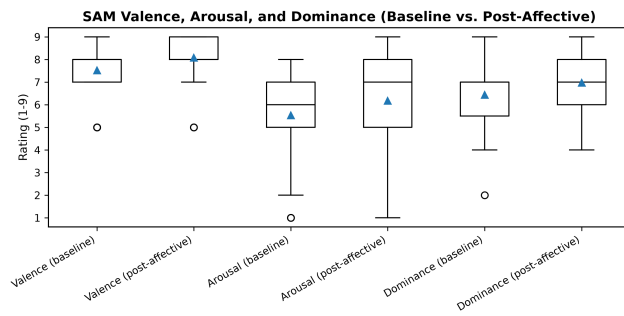
- [41] Olga Sutsikova, Atsushi Senju, and Tim J. Smith. 2023. Cognitive Impact of Social Virtual Reality: Audience and Mere Presence Effect of Virtual Companions. *Human Behavior and Emerging Technologies* 2023, 1 (2023), 6677789. doi:10.1155/2023/6677789 _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1155/2023/6677789>.
- [42] V. Vinayagamoorthy, M. Gillies, A. Steed, E. Tanguy, X. Pan, C. Loscos, and M. Slater. 2006. *Building Expression into Virtual Characters*. Technical Report. Eurographics Association.
- [43] David Watson, Lee Anna Clark, and Auke Tellegen. 1988. Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology* 54, 6 (1988), 1063–1070. doi:10.1037/0022-3514.54.6.1063
- [44] Carolin Wienrich, K. Schindler, Nina Döllinger, S. Kock, and O. Traupe. 2018. *Social Presence and Cooperation in Large-Scale Multi-User Virtual Reality - The Relevance of Social Interdependence for Location-Based Environments*. doi:10.1109/VR.2018.8446575 Pages: 214.
- [45] Jason W. Woodworth and Christoph W. Borst. 2024. Design and Validation of a Library of Active Affective Tasks for Emotion Elicitation in VR. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. 398–407. doi:10.1109/VR58804.2024.00061 ISSN: 2642-5254.
- [46] Shuran Yang and Wenxiang Zhang. 2022. Presence and Flow in the Context of Virtual Reality Storytelling: What Influences Enjoyment in Virtual Environments? *Cyberpsychology, Behavior, and Social Networking* 25, 2 (Feb. 2022), 101–109. doi:10.1089/cyber.2021.0037
- [47] N.L. Yeo, M.P. White, I. Alcock, R. Garside, S.G. Dean, A.J. Smalley, and B. Gatersleben. 2020. What is the best way of delivering virtual nature for improving mood? An experimental comparison of high definition TV, 360° video, and computer generated virtual reality. *Journal of Environmental Psychology* 72 (Dec. 2020), 101500. doi:10.1016/j.jenvp.2020.101500

A Supplementary Result Figures

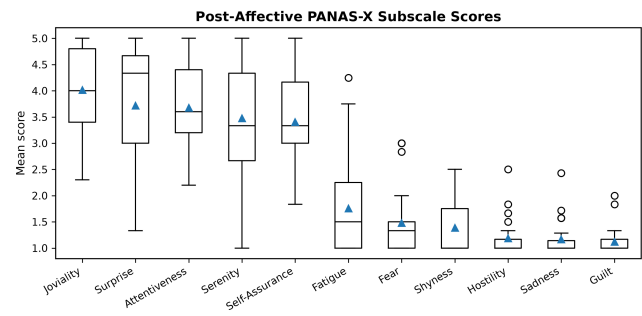
Figures 7, 8, and 9 present the SAM, PANAS-X, ANSS, behavioural engagement, temporal engagement, and outlier-robustness results referenced in Section 5 graphically.

B Convergent Evidence Summary

Table 4 summarises the outcome against each of the five elicitation criteria defined in Section 4.5, alongside the specific evidence that qualifies or counts against each one; a criterion being “met” in the left-hand sense does not mean the evidence for it is uniformly strong, and the right-hand column should be read as an integral part of the result, not a footnote to it.

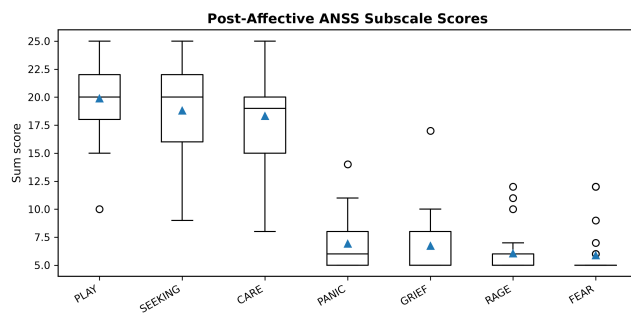


(a) SAM valence, arousal, and dominance.

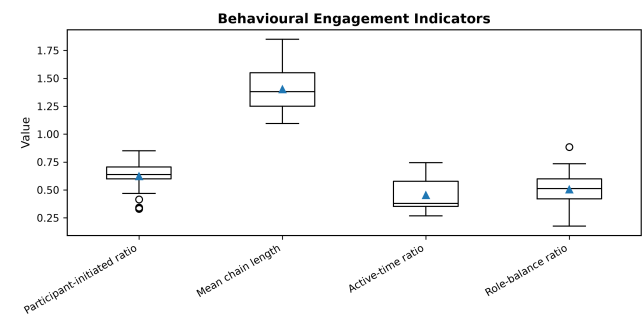


(b) Post-affective PANAS-X subscale distributions.

Figure 7: Baseline, post-tutorial, and post-affective SAM distributions, and post-affective PANAS-X subscale distributions.

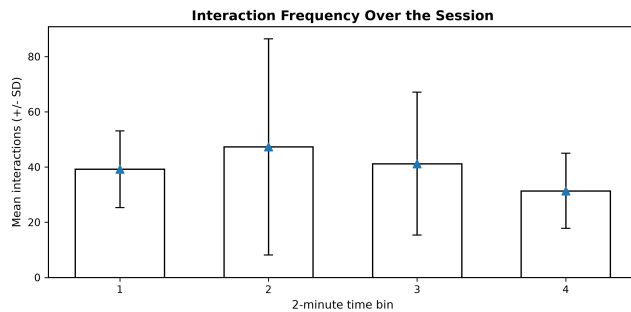


(a) Post-affective ANSS subscale distributions, showing PLAY relative to competing subscales.

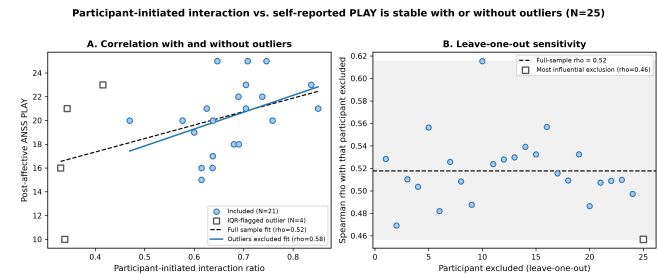


(b) Behavioural engagement indicators: participant-initiated ratio, mean chain length, active-time ratio, and role-balance ratio.

Figure 8: Post-affective ANSS distributions and behavioural engagement indicators.



(a) Interaction frequency across the four two-minute bins of the session.



(b) Robustness of the participant-initiated ratio vs. ANSS PLAY correlation: outlier-marked scatter with regression lines (A) and leave-one-out sensitivity (B).

Figure 9: Temporal engagement and outlier-robustness check for the participant-initiated interaction correlation.

Table 4: Summary of evidence for and against PLAY elicitation.

Criterion	Met?	Evidence for	Evidence against / caveats
ANSS PLAY (Pankseppian, state-level) significantly higher than next-highest subscale, and rises across the three stages	Yes	PLAY > SEEKING post-affectively ($p = .017$); significant 3-stage change ($p < .001$)	No corrected pairwise contrasts available to attribute the 3-stage rise specifically to the Apple VE rather than to entering VR generally
SAM VAD profile matches the target high-VAD region	Partial	Valence and dominance above threshold; valence rises significantly from true baseline	Arousal does not exceed threshold; dominance shows no reliable change once a true pre-VR baseline is used (3-stage omnibus ns)
Presence significantly above reference point	Yes	IPQ > 0 post-affectively ($p < .001$)	IPQ does not increase further from post-tutorial to post-affective; presence was established by the tutorial alone
Co-presence significantly above midpoint	Yes	$t(24) = 4.56, p < .001$	No pre-affective (tutorial) baseline exists for co-presence, so a general-VR-agent effect cannot be ruled out
Social presence significantly above midpoint	No	—	Mean not distinguishable from ambivalence ($p = .691$)
Engagement sustained (no decline) over time	Yes	Friedman ns across 4 bins ($p = .449$)	Absence of decline is not itself evidence of high engagement, only of no fatigue-driven drop-off
At least one behavioural indicator correlates with subjective PLAY	Yes	Participant-initiated ratio, $\rho = .52, p = .008$; robust to leave-one-out and outlier removal (Appendix A)	Five of six indicators tested against ANSS PLAY were not significant; the one significant indicator is correlational, not causal

Table 5: Three-stage results, remaining PANAS-X subscales.

Subscale	Pre	Post-tut	Post-aff	Omnibus
Hostility	1.23 (0.33)	1.21 (0.35)	1.19 (0.36)	ns ($p = .624$)
Attentiveness	3.42 (0.79)	3.74 (0.72)	3.68 (0.85)	$p = .037$
Shyness	1.86 (0.53)	1.56 (0.57)	1.39 (0.53)	$p = .002$
Fatigue	2.56 (1.05)	1.99 (0.87)	1.76 (0.95)	$p < .001$
Self-Assurance	3.15 (0.75)	3.23 (0.91)	3.41 (0.89)	ns ($p = .127$)
Serenity	3.64 (0.86)	3.56 (0.82)	3.48 (1.12)	ns ($p = .753$)
Guilt	1.42 (0.50)	1.22 (0.42)	1.12 (0.26)	$p < .001$
Sadness	1.42 (0.49)	1.22 (0.37)	1.17 (0.33)	$p = .009$
Fear	1.81 (0.86)	1.51 (0.67)	1.48 (0.62)	$p = .020^a$

Mean (SD) at each stage. ^aSignificant omnibus test, but no individual pairwise comparison survived Holm-Bonferroni correction.

Table 6: Three-stage results, remaining ANSS subscales.

Subscale	Pre	Post-tut	Post-aff	Omnibus
SEEKING	17.56 (3.34)	19.48 (3.42)	18.80 (4.29)	$p = .019^a$
CARE	17.80 (3.43)	17.68 (4.02)	18.32 (4.50)	ns ($p = .451$)
PANIC	8.52 (3.94)	7.24 (3.22)	6.92 (2.43)	$p = .009$
GRIEF	6.72 (2.05)	7.04 (2.94)	6.72 (2.78)	ns ($p = .355$)
FEAR	6.44 (2.42)	5.68 (1.44)	5.88 (2.05)	ns ($p = .461$)
RAGE	6.80 (3.07)	6.00 (1.68)	6.04 (2.01)	ns ($p = .273$)

Mean (SD) at each stage. ^aSignificant repeated-measures ANOVA; no corrected pairwise contrasts available (see Section 5.2 for the same caveat on ANSS PLAY).

C Additional Statistical Results

This appendix reports statistical results referenced in Section 5 but not required there to evaluate the paper’s primary claims: the full three-stage (pre-tutorial / post-tutorial / post-affective) results for every PANAS-X and ANSS subscale not already reported in Table 3, and the full behavioural-subjective correlation matrix underlying the single correlation reported in Section 5.4.

Table 5 and Table 6 report the remaining PANAS-X and ANSS subscales respectively; Figure 11 shows the same data visually, as a three-stage trajectory per subscale, so the shape of each change

(rising, falling, or flat) can be seen at a glance alongside the exact numbers in the tables. Three PANAS-X subscales showed no significant omnibus change across the three stages: hostility, self-assurance, and serenity. Of the remaining six PANAS-X subscales, all showed a significant omnibus change; fear is a notable case where the omnibus test was significant ($\chi^2(2) = 7.83, p = .020$) but none of the three individual pairwise comparisons survived the Holm-Bonferroni correction, illustrating why the two-step procedure in Section 4.5 (only running pairwise tests after a significant omnibus test) is not equivalent to guaranteeing an interpretable pairwise result. Among the remaining ANSS subscales, SEEKING and PANIC showed significant omnibus change while CARE, GRIEF, FEAR, and RAGE did not; PANIC’s pattern, decreasing from pre-tutorial to both later stages but not changing further between post-tutorial and post-affective, is discussed in Section 6 as a plausible, though ad hoc, indicator of general anxiety reduction on entering VR rather than a PLAY-specific effect.

Figure 10 shows the full correlation matrix between all six behavioural indicators and all six post-affective subjective outcomes (36 tests in total), underlying the single significant result (participant-initiated ratio vs. ANSS PLAY) reported in Section 5.4. At an uncorrected $\alpha = .05$, roughly 1.8 significant cells would be expected by chance alone across 36 independent tests; two cells reach significance, the previously reported participant-initiated ratio vs. ANSS PLAY ($p = .52, p = .008$), and time-in-proximity ratio vs. SAM valence ($p = -.43, p = .032$). The latter was not hypothesised in advance, involves the one behavioural variable already flagged in Section 4.4 as having a likely logging-threshold problem, and is consistent in count with what chance alone would predict across this many tests; it is reported here for completeness and transparency but is not treated as evidence of a real effect, and is not relied upon anywhere else in this paper

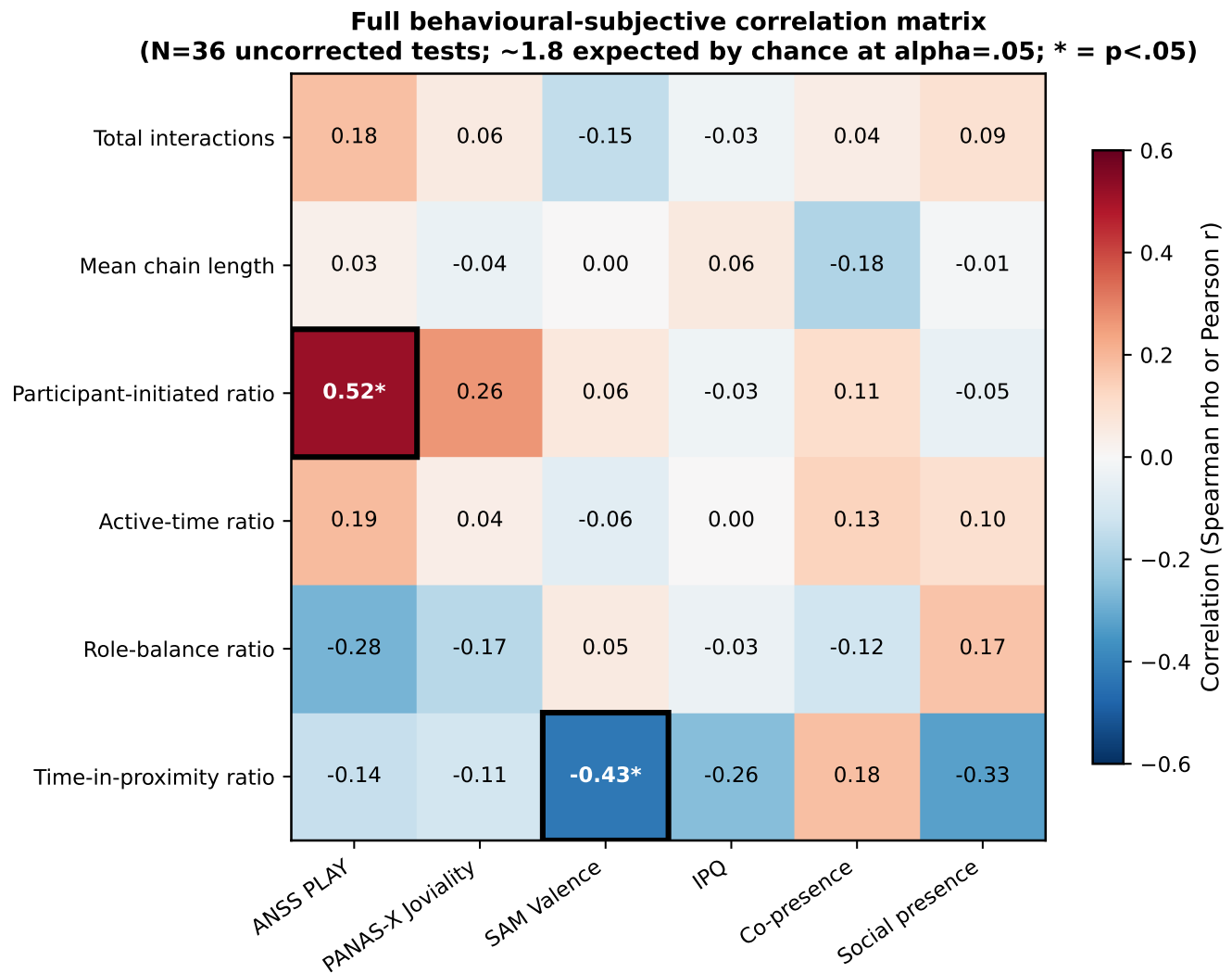


Figure 10: Full behavioural-subjective correlation matrix (Spearman ρ or Pearson r , method chosen per pair by the same normality criterion as Section 4.5). Bordered cells are significant at uncorrected $p < .05$.

**Three-stage trajectories for secondary PANAS-X and ANSS subscales
(blue/filled = significant omnibus test, grey/open = not significant)**

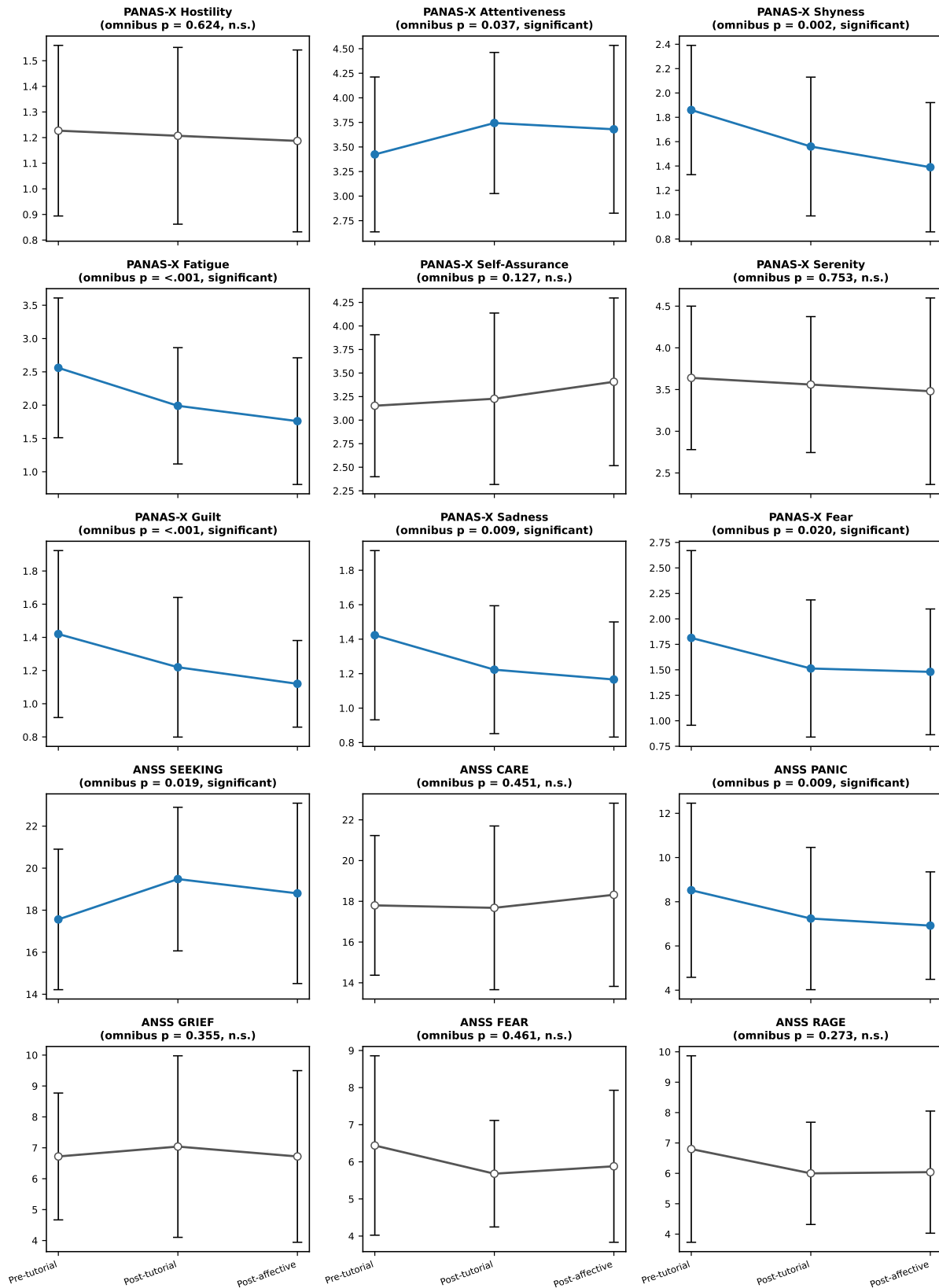


Figure 11: Three-stage trajectories (mean $\pm$ SD) for the PANAS-X and ANSS subscales in Tables 5 and 6. Filled blue points and lines indicate a significant omnibus test; open grey points and lines indicate no significant change.