

Breaking The Experience: Effects of Questionnaires in VR User Studies

Susanne Putze^{*,1}, Dmitry Alexandrovsky^{*,1}, Felix Putze²,
Sebastian Höffner¹, Jan David Smeddinck³, Rainer Malaka¹

¹ Digital Media Lab, University of Bremen, Germany

² Cognitive Systems Lab, University of Bremen, Germany

³ Open Lab, School of Computing, Newcastle University, UK

{sputze,dimi,fputze,shoeffner,malaka}@uni-bremen.de Jan.Smeddinck@newcastle.ac.uk

ABSTRACT

Questionnaires are among the most common research tools in virtual reality (VR) evaluations and user studies. However, transitioning from virtual worlds to the physical world to respond to VR experience questionnaires can potentially lead to systematic biases. Administering questionnaires in VR (INVRQs) is becoming more common in contemporary research. This is based on the intuitive notion that INVRQs may ease participation, reduce the Break in Presence (BIP) and avoid biases. In this paper, we perform a systematic investigation into the effects of interrupting the VR experience through questionnaires using physiological data as a continuous and objective measure of presence. In a user study ($n=50$), we evaluated question-asking procedures using a VR shooter with two different levels of immersion. The users rated their player experience with a questionnaire either inside or outside of VR. Our results indicate a reduced BIP for the employed INVRQ without affecting the self-reported player experience.

Author Keywords

Virtual reality; VR; user studies; in-VR questionnaires; inVRQs; break in presence; surveys; biosignals; research methods.

CCS Concepts

•Human-centered computing → Virtual reality; HCI design and evaluation methods; Empirical studies in HCI; User studies;

INTRODUCTION

Recent advances of VR technology have enabled new research methods and interventions across various fields and allow for the design of highly immersive applications that evoke a strong sense of presence. Subjective responses through questionnaires remain a widely applied method for administering

*The authors contributed equally.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHI '20, April 25–30, 2020, Honolulu, HI, USA.

Copyright is held by the owner/authors(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-6708-0/20/04 ..\$15.00.

<http://dx.doi.org/10.1145/3313831.3376144>

mid- and post-experience measures [52]. Research with – as well as the development and evaluation of – VR experiences often relies on questionnaires which are completed on paper or through computer-based forms. Thereby, participants are required to leave the VR, i.e. take off the HMD, to return to the physical domain [44] and experience a break in presence (BIP) [89]. The BIP is associated with physiological effects [85, 86], disorientation and loss of sense of control [48]. Accordingly, questionnaire results are likely biased to a degree that is difficult to quantify and may vary across individuals. Such undetermined bias is highly problematic for many types of research that are based on subjective measurements.

In contrast to the physical domain, alternate reality technologies allow for the embedding of questionnaires directly into the virtual environment (VE). Embedding question-items in the VE offers an opportunity to stay closer to the context of an ongoing experience [32] than out-VR data acquisition allows, as it avoids a drastic change of context – BIP [3, 44, 74]. Schwind et al. have shown that administering questionnaires directly in VR contributes positively to the consistency of self-reports on presence. They advocate for surveying participants directly in the VE [74]. However, these results were only validated for self-reports on presence and might differ for other constructs, especially since presence is a primary candidate for being affected by BIP. The publication also does not report measured BIP in relation to the outcomes.

Recent research has already started to embed questionnaires in the VE for various applications [45, 63, 76]. However, related work does not offer considerations about the extent and influence of BIPs on the VR experience, as triggered by posing questionnaires [3]. It appears crucial to investigate the side-effects of question-asking in VR user studies as researchers have to be aware of biases that may exist for their research methods of choice. In this paper, we address this gap and investigate the effects of BIPs evoked through question-asking in VR via a laboratory user study. We imitate a representative VR user study with repeated self-report measures and administer questionnaires both embedded into the VE (INVRQ) and outside of the VE (OUTVRQ). For our investigation we developed the following hypotheses:

H₁: *Switching from virtual experiences to completing questionnaires produces a physiologically measurable BIP.*

H₂: *There is a measurable difference of BIP between the administration of INVRQs and OUTVRQs.*

H₃: *Completing questionnaires outside VR and in VR has a measurable effect on performance in successive trials.*

H₄: *There are measurable (uncontrolled) biases in self-reports triggered by BIP between INVRQs and OUTVRQs.*

We conducted a mixed-design user study ($n=50$) and employed physiological measures of BIP as well as subjective ratings of presence and player experience (PX). Our results show a clear difference of BIPs between INVRQs and OUTVRQs and support the evidence from prior research [31, 74, 89, 85] that less-invasive methods of self-reports increase the validity of the measures. Our findings highlight the effects of self-reporting in VR and help both researchers and developers to consider them in their study design.

RELATED WORK

The sense of *presence* is considered to be the key feature of VR [23, 41] and is often treated as a metric of effectiveness of VEs [67, 102]. It is commonly defined as the subjective experience of “being there” [60] – in the virtual environment rather than in the physical space [25, 81, 84, 102]. Similarly, Lombard and Ditton identify presence as the disappearance of the communication medium – a “perceptual illusion of nonmediation” [55], which is consistent with Bystrom et al.’s suspension of disbelief [13]. Slater et al. extend this definition and describe presence as “how well a person’s behaviour in the VE matches their behaviour in similar circumstances in real life” [88] and argue that how data is displayed – and how the participants are able to interact in VR – is more important for presence than the level of realism [67, 84]. Zahorik and Jenison also claim that plausible responses of the VE to the user’s actions engage the sense of presence [103]. IJsselsteijn et al. adapt Lombard’s and Ditton’s definition and distinguish two subcategories of physical and social presence. *Physical presence* refers to the sense of a physical location and *social presence* describes the feeling of “being together” [42]. The authors also provide four determinants of presence: extent and fidelity of sensory information, match between sensors and display, content factors and user characteristics. Likewise, Sheldrian proposed three determinants of presence: extent of sensory information, control of relation of sensors to environment and ability to modify the physical environment [78] and Witmer and Singer defined four factors that contribute to the sense of presence: control, sensory, distraction and realism [102]. The immersion, presence, performance model [13] follows Slater’s definition [88] and proposes a feedback loop between attention, engagement and presence which also underlies the flow theory [18].

While there are diverse definitions of presence across the literature (cf. [38, 55, 82, 102]), there is a consensus that presence is a multidimensional construct that is driven by media characteristics, such as technological factors [99, 56, 68, 81], and personal characteristics [4]. Skarbez et al. [81] provide a comprehensive review of literature on presence. The authors identify similarities between the definitions and aggregate common variables and related constructs that contribute to the

sense of presence into a conceptual model of *Place Illusion*, *Plausibility Illusion* and *Social Presence Illusion*.

In contrast to presence, *immersion* is recurrently described as the objective properties of the VE and the applied technology that induce the sense of presence [9, 88, 92]. Immersion and presence are two logically separated constructs, however they are directly related [84] as presence is the outcome of immersion [70]. Brown and Cairns [12] identified *engagement*, *engrossment* and *total immersion* as three levels of immersion in games on a scale of involvement. Immersion includes the software and the hardware components that produce stimuli to the user’s senses and affect how the user perceives the VE [65]. Stereoscopic rendering, resolution, frame rate, field of view, levels of user tracking and fidelity of sensory input are considered as driving factors that enable immersion and thus, facilitate the sense of presence [9, 19, 36, 39, 58, 91, 99]. From a content analysis of 83 studies, Cummings and Bailenson infer that technological immersion has a medium-sized effect on presence [19]. Witmer and Singer conclude that both involvement and immersion are vital to invoke the sense of presence [102].

The relationship between presence and performance is ambiguous. While some literature showed a positive correlation between presence and performance [80, 88], others could not find a significant relationship [61, 47, 104]. Welch argues that presence does not facilitate performance, arguing that positive effects are most likely due to the increased immersive properties of the VE (e.g. frame rate, latency, resolution) [98]. Similarly, Bystrom et al. contend, that the relationship between presence and performance is task-specific and performance only increases with greater presence if the latter is relevant for the task [13]. However, a substantial body of research shows evidence that the immersive characteristics of a VE can promote task performance [62, 71, 94], learning outcomes [1, 16, 17, 57] and therapy effects [15, 23, 41, 31].

Measuring Presence

Self-reports

Post-experience presence questionnaires are the predominant method applied in the literature [81, 89]. Subjective measurements of presence should be relevant, sensitive, convenient, non-intrusive and reliable [39]. We are aware that a significant body of work developed questionnaires to assess subjective sense of presence. However, in this review we focus only on the three most commonly used questionnaires in contemporary research [74]. Witmer and Singer developed a 32-item *presence questionnaire* (PQ) with the subscales Involvement/Control, Natural, Auditory, Haptic, Resolution and Interface Quality [102]. The *Slater-Usoh-Steed* (SUS) questionnaire consists of 6 items [91] on a 7-point Likert scale with no separate subscales. The authors argue that any measuring instrument of presence needs to be validated against responses in a real environment. To verify their questionnaire, the authors contrasted the SUS and the PQ between a task in VE and real environment. The results show marginal differences between real and virtual results on the SUS but not on the PQ [96]. Based on items from previous work [15, 26, 40, 65, 91, 95, 102], Schubert et al. developed the *igroup*

presence questionnaire (IPQ) which consists of 14 items on a 5-point Likert scale with the subscales *General Presence*, *Spatial Presence*, *Involvement* and *Realism* [70].

The major advantage of questionnaires is that they are easy to administer and generally don't require modifications of the VE [81]. However, since questionnaires on presence are most commonly conducted post-experience, they make for intrusive and not continuous experiences. Therefore, they are inherently unreliable for assessing presence [31, 81, 83] and they are not sensitive to state changes during the ongoing experience [54, 85]. Apart from questionnaires, research proposed relative measures using a number rating [99], a continuous scale [93] or physical sliders [31] for perceived presence. Nonetheless, these assessment methods share the same deficiency as questionnaires [81]. Accordingly, Freeman et al. suggest the application of objective measurements of presence [31].

Behavioral Assessment

Several approaches have been proposed for behavioral measures of presence, including responses to social or threatening stimuli [78] (e.g. ducking from a flying to object), measures of discrepancy between stimuli inside and outside the VE [90], or magnitude of postural responses [30]. The rationale behind this approach is that a higher degree of presence should result in stronger behavioral responses to the stimuli in the VE [30]. Skarbez et al. point out that behavioral measures are objective, contemporaneous and non-intrusive and thus, they overcome some of the shortcomings of the subjective measures. However, in order to trigger specific behavioral responses the VE and any ongoing study or evaluation procedure requires specific manipulations, which are not always applicable [81].

Physiological Measures

Physiological responses provide information about specific episodes of the experience [54] and allow a better interpretation of subjective ratings and task performance [11]. Highly immersive experiences are expected to facilitate specific reaction patterns from the autonomous nervous system [20]. Meehan reported a significant correlation between self-reported presence and skin conductance in a stressful VE, while neither skin temperature, nor heart rate (HR) [58, 59] showed significant effects. Slater et al. used the same (stressfull) VE to investigate the effects of visual realism on presence. The authors found a significant increase of electrodermal activity (EDA) and HR over baseline for the high fidelity variant, while the low fidelity VE didn't differ from baseline [87]. Dillon et al. correlated EDA and electrocardiogram (ECG) signals to the ITC-SOPI [53] presence questionnaires and could not find a correlation between physiological signals and subjective self-reports on presence [21, 20]. In contrast, Wiederhold et al. found significant negative correlations between subjective self-reports of presence/realism and EDA/HR. Contrary to their expectation, HR was also negatively correlated with realism and presence [100]. Brogni et al. found that physiological stress increases in immersive VEs and decreases in natural-looking and engaging VEs [10, 11]. Baumgartner et al. attributed specific activity of the dorsolateral prefrontal cortex and dorsal stream in fMRI signals modulating the ex-

perience of presence in adult subjects [5]. These regions are responsive to control of attention, orientation and control of egocentric orientation [43] which is in line with many definitions of presence discussed above. Similarly, Bouchard et al. identified activation of the parahippocampus regions corresponding to the sense of presence [7]. Skarbez et al. conclude that physiological measures are "truly objective", since they are contemporaneous and continuous. Yet again, they are also cumbersome to administer [81]. Slater et al. argue that physiological measures of presence can only be applied in anxious scenarios (e.g. a response to a threat) but that they are ineffective in mundane situations [84].

Break in Presence

An alternative approach to measuring presence is based on the assessment of "breaks in presence" (BIPs) [89]. BIP describes the moment when the illusion generated by the VE collapses and the user switches awareness from the VE to the physical environment [85], or when the users experience inconsistencies between their mental model and the VE [54]. BIPs can be caused, e.g. by loss of tracking, glitches, or noises outside the VE [44] and may provoke negative emotions [69]. Slater and Steed proposed a stochastic model of presence over time that relies on self-reports of transitions between "presence in the VE" to "presence in the real world", rather than on post-experience assessment of presence. Based on BIP, a Markov chain was induced that continuously modeled the user's state of presence. In contrast to other self-report measures e.g. questionnaires, this method assesses presence after it has been disturbed and thus is minimally invasive [67, 81, 83]. Further, it allows for investigating the cause of a particular BIP, since the moment of the BIP can be determined precisely [89]. A BIP causes a (moderately) shocking experience which invokes a physiological reaction [85] similar to the responses in physiological measures of presence, cf. [58]. Slater et al. examined the physiological response to BIPs in six different VEs and found changes in EDA and HR as indicators of BIP [85]. Liebold et al. used ECG, EDA and muscle activity (EMG) together with behavioral data measures to differentiate between ten most frequent and intense types of BIPs in a commercial PC game. The results show that gameplay interruptions produce the strongest EDA and thus the most intense BIP [54], an effect Slater et al. observed earlier when measuring HR, HRV and EDA as indicators of BIP in a "sudden whiteout" scenario [86].

Question-asking in VR

Recent research started taking interest in administering questionnaires in VR. Schwind et al. investigated the effect of filling out questionnaires on presence in VR. 36 participants played a VR shooter with two degrees of realism and filled out 3 presence questionnaires (IPQ [70], SUS [96], and PQ [102]) on a desktop PC or on a virtual PC in a replicated lab in VR. The authors found no differences on the mean scores, but the responses in VR showed a significantly lower variance and therefore, higher consistency of the results [74]. These results support the aim of finding noninvasive measures of presence [31, 93, 99, 89]. Alexandrovsky and Putze et al. examined the contemporary usage of INVRQs and evaluated

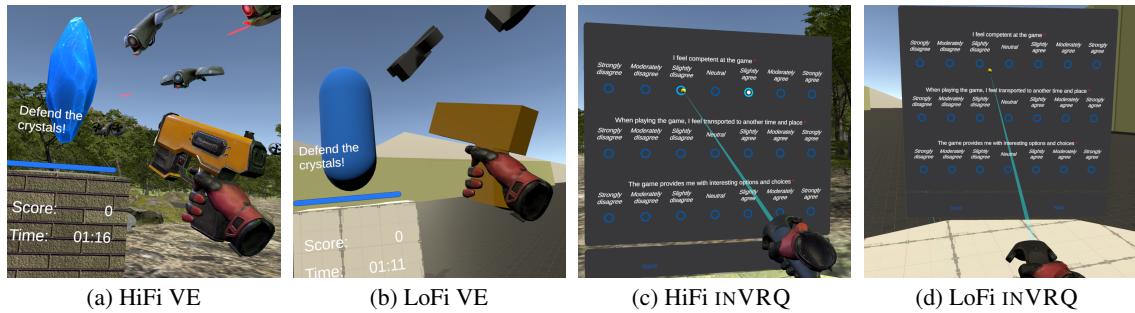


Figure 1. HiFi and LoFi variants of the virtual environment and the INVRQs.

the design of INVRQs in two user studies. The authors could confirm Schwind et al.’s comparable means in and out VR but could not find higher consistency of INVRQs. Further, they report slightly lower—but still high—usability rating for INVRQs in contrast to OUTVRQs with higher user enjoyment of INVRQs [3]. Kang et al. [45] included a single item full screen overlay (i.e. HUD) between multiple trials in their VR environment. Participants interacted with a native VR controller to select between two or more unordered answers [45]. Schwind et al. [73, 75, 76] presented the full 32-item PQ [102] on a 3D floating UI showing one-line text instructions and 4 items on Likert-scales. Users selected answers and navigated through the questionnaire with freehand gestures. Oberdörfer et al. [63] presented a paper-like, world-referenced NASA-TLX [37] in VR. The subjects interacted with the questionnaire using a native VR controller as a pointer. Wienrich et al. [101] applied a body-referenced INVRQ in their user study. The questionnaire is displayed on a 2D floating UI with a 20-items PANAS [50] attached to the hand of the virtual character. Fernandes and Feiner [29] applied a 10-point slider in VR as a measure of discomfort; this method was further adapted in recent research [2, 14, 35].

All INVRQ designs differ in their presentation (2D overlay, world-reference, body-referenced), their extent (single-question vs. multi-item questionnaire), question-item presentation (text-based vs. scales) and interaction modalities (pointing, free-hand, gamepad). These differences highlight a lack of standardized tools to assess self-reports in VR.

Background Summary

Previous research points out that presence is a crucial aspect of VR [23, 41, 67, 102]. Further, a break in presence threatens the VR experience and should be minimized [44]. As the ongoing debate on questionnaires for measuring presence highlights, they likely cause a break in presence [89, 81] and such sudden interruptions or transitions between realities can affect the emotional state negatively [48]. Accordingly, assessing presence within immersive experiences using biosignals suggests more reliable measurements [8, 31, 89]. However, although research has shown that INVRQs can provide more consistent results [74] for measures of presence, related work does not offer insights about the degree of BIP caused by filling out a questionnaire and thus offers limited insights into potential systematic influences, which would be valuable insights, since

any such systematic influences could statistically be accounted for.

STUDY

In our study, we focus on BIP caused by discontinuation of the VR experience to fill out questionnaires. Therefore, in alignment with Schwind et al. [74], we created an immersive VE in which participants were required to engage with a playful tasks repeatedly at different levels of realism and respond to questionnaires inside and outside VR. We recorded physiological signals during the whole session, since they are more sensitive to assess BIPs than subjective self-reports [89].

The study design employs mixed methods and contains two independent variables: questionnaire modality and fidelity. The questionnaire modality describes the presentation of a questionnaire, either as INVRQ or OUTVRQ. Since physiological responses are highly specific and can vary drastically between participants we chose a within-subjects design with repeated measures for the questionnaire modality. This also allows the participants to compare both assessment methods. The order of the questionnaire conditions was randomized. Fidelity is operationalized as degree of immersion [10, 46, 74] with the levels low (LoFi) and high (HiFi). Based on prior research which showed evidence that visual fidelity fosters immersion and therefore affects the sense of presence [9, 87, 104], we hypothesize that a switch from LoFi to physical reality would cause a smaller BIP compared to switching from HiFi. To avoid transfer effects, fidelity was administered between-subjects.

The Virtual Environment

In the VE, the player is located on a plateau in open space surrounded by three crystals which are attacked by drone enemies. The task of the game is to shoot the drones with a pistol using the VR controller. To eliminate a drone, the player is required to hit it twice. This is to decelerate the body movement of the player and thus reduce artefacts in the biosignals. We aimed to provide a sustained medium-intensity engagement with the game; players should feel connected to the task while not being overly aroused or stressed. Therefore, we balanced the game so that the player would lose the round within the last seconds if they did not eliminate any drones. Since we attached biosensors to the non-dominant hand, we deliberately designed the game to be playable with one hand only. To operationalize different levels of presence, we altered the visual and aural fidelity of the VE. For the HiFi variant (stylized), we

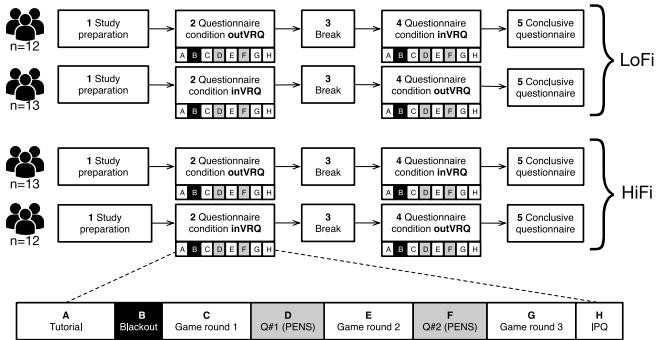


Figure 2. Flowchart of the study procedure with questionnaire modality (inVRQ, outVRQ) as a within-subject variable in randomized order and fidelity (HiFi, LoFi) as between-subjects conditions. States 2 and 4 in each flow diagram consist of the steps A-H. Questionnaires in states D and F are conducted in or out of VR in the respective condition. Black and grey states (B, D, F) indicate the investigated BIP events. The IPQ at the end of each condition is assessed only outside VR.

used high resolution textures, sound, physics simulation and particle effects (Fig. 1a). The LoFi environment (abstract) only consisted of primitive mesh objects which approximated the HiFi objects without particle effects nor sounds (Fig. 1b). We payed attention to design both fidelity variants with comparable difficulty and only altered the visuals and sounds with the same hidden models for collision-detection. To avoid learning effects as confounds in the biosignals and performance, the drones spawned randomly at two fixed locations. Besides the altered fidelity and the randomized spawn-points, there were no differences in the VE between the trials. The environment was implemented in Unity3D and run on a high-end PC with a HTC Vive Pro HMD at a constant frame rate of 90 FPS.

inVRQ Design

We employed the INVRQ tool, which was previously designed and evaluated by Alexandrovsky and Putze et al. [3]. The design follows general guidelines from traditional UIs [24, 79] and VR interface design [22, 34, 33, 64] and received high usability scores [3]. The questionnaire is anchored in world space and users interact with a controller using a common laser pointer metaphor (cf. Fig. 1c,1d). We kept INVRQs constant and in the same position in both variants of the VE.

Recording of Physiological Signals

We used a Mind Media NeXus 10 MKII biofeedback device¹ with the BioTrace+ V2018A software for recording of the physiological signals with a sampling rate of 128Hz. The NeXus 10 was connected using a 5m USB cable allowing the participants to move around freely in the tracking space. We used a chest strap respiration sensor, a blood volume pressure (BVP) sensor and a skin conductance (SC) sensor to measure electrodermal activity (EDA) on the non-dominant hand of the participant. These biosignals are in alignment with physiological measures of presence [53, 58, 59] and BIP [54, 85, 86].

Figure 3 illustrates the recording setup: We attached the SC sensors using adjustable velcro straps to the middle phalanx

¹<https://www.mindmedia.com/en/products/nexus-10-mkii/>



Figure 3. A user wearing the HMD and biosensors.

of ring and little finger which have the highest SC responsiveness [77, 97]. To synchronize the recordings of the signals with the game, we used audio signals and manual triggers. Before placement of the electrodes, the participants cleaned their non-dominant hand with a wet wipe. To get a clear signal quality and reduce artifacts due to movement, we briefed the participants not to use their non-dominant hand with the sensors and to let it hang down during the whole study. A conductor helped the participants with fitting the HMD.

Subjective Measures

To assess player experience after each game round, we applied the Player Experience of Need of Satisfaction (PENS) [66] questionnaire either using an INVRQ or OUTVRQ. It consists of 21 items on a 7-point Likert scale with the subscales of autonomy, competence, relatedness, presence, and intuitive control. With 21 items PENS is similar in length to other questionnaires (e.g. PANAS, IPQ, NASA-TLX) used in previous user studies with INVRQs [45, 63, 76]. To assess potential differences in perceived sense of presence due to different questionnaire modalities and to validate the “breakable experience”, we operated the igroup presence questionnaire (IPQ) [70] after the third game rounds in each condition (2× for each participant). The IPQ consists of 14 items on a 5-point Likert scale with the subscales General Presence (GP), Spatial Presence (SP), Involvement (INV) and Realism (REAL). We did not assess the IPQ between trials because it is not sensitive for measuring a BIP [3, 74] and the game rounds did not differ. We also collected self reports about game experience and usability with both questionnaire modalities. Finally, the users ranked the BIP events by the degree of distraction.

Procedure and Tasks

Our study flow is depicted in Figure 2 and consisted of the following states: 1. Study preparation: Briefing and complete consent form. Random assignment to a condition (HiFi or LoFi) and order of the questionnaire modality. Attach physiological sensors and synchronize biosignals with the game. Put on the HMD. 2. First questionnaire condition, steps A.-H. (INVRQ or OUTVRQ). 3. Break (optional). 4. Second questionnaire condition, steps A.-H. (OUTVRQ or INVRQ).

5. Conclusive questionnaire on a PC with demographics, ranking questionnaire modalities and debriefing.

Each questionnaire condition contained the following steps: A. Put on the HMD and play a tutorial round (60 s). B. Initial blackout BIP. C. Game round #1 (90 s). D. PENS #1 using INVRQ or OUTVRQ depending on the condition. E. Game round #2 (90 s). F. PENS #2 using same questionnaire modality as PENS #1. G. Game round #3 (90 s). H. Take off the HMD and fill out an IPQ on the PC. The whole procedure required around 45 min with an average in-VR time of 17.35 min ($SD=1.45$). The study took place in a lab room without irregular light, climate or noise conditions.

Participants

53 participants, mostly students from computer science and related areas from the local university, volunteered to participate in our study. All participants qualified for an Amazon voucher lottery. Due to technical problems we excluded three participants from the analysis. In the following, we consider $n=50$ participants (8 female [self-identified], mean age $M=26.08$, $SD=4.39$). One participant suffers from red-green color blindness, 45 participants are right-handed, 3 left-handed and 2 bi-manual, who decided to interact with their right hand. 20 participants used vision correction during the study. The four participant groups were balanced for gender (cf. Figure 2). The mean game experience was $M=2.76$, $SD=2.17$ (1–8 scale, 1 max) and the mean VR experience was $M=6.48$, $SD=2.22$ (1–8 scale, 1 max). One-way ANOVAs showed no significant differences across conditions for both.

Pre-processing

We used Python 3.7 and R with a Jupyter notebook for the data analysis. The recording software BioTrace+ V2018A features pre-processing of the biosignals: From the BVP sensor, it calculates heart rate (HR) and the heart rate variability (HRV), from the signals of the respiration sensor it calculates respiration rate (RR). The signals were exported in raw format from the recording software with a sampling rate of 128 Hz. The HR and the EDA signal were cleaned up using a Butterworth lowpass filter of order 5.

RESULTS

In the following section we focus on significant results regarding our hypotheses.²

Physiological Signals

In this subsection, we inspect two aspects of the physiological signals: First, we compare the magnitudes of physiological response to the INVRQ and OUTVRQ events with the signals resulting from the blackout event. Second, we analyze the influence of an INVRQ or OUTVRQ on the subsequent tasks to study any potential physiological aftereffects following such report-related BIPs.

Detection and Comparison of BIP events

In alignment with work by Liebold et al. [54] and Slater et al. [86], we employed 3 second blackouts during a first tutorial

²The (complete) analysis and data are available at open science framework: <https://osf.io/cgsqa>

round of the VR game as a reference BIP event, which we compared to the INVRQ and OUTVRQ events. We expect a reaction in the biosignals to the blackout and the INVRQ event directly after the event. The analysis of OUTVRQ events is less straightforward: At $t=0$ s the participants were asked to contact the experimenter for taking off the HMD, requiring $M=13.78$ s ($SD=3.85$). Thus, we analyzed both post-BIP events, the occurrence of the “contact the experimenter” (red line, Fig. 4c) notification and the HMD-off event (black line). Figure 4 shows the physiological response to the blackouts (Fig. 4a), the the INVRQs (Fig. 4b) and OUTVRQs (Fig. 4c), each averaged over all participants. For EDA, we followed the procedure of Liebold et al. [54] and analyzed the phasic skin conductance response [6] as an indicator for event-related activity. We excluded the signals of one participant from the EDA analysis due to an unusually high baseline signal.

For every BIP event, we compared the measure from a time interval of seconds 3 to 6 after the event with a 3 s baseline window just before the task using paired samples t-tests. The intervals investigated are illustrated in Figure 4 with dashed green lines. As indicated in Figure 2, each questionnaire condition contains three different BIP events, leading to a total of six BIP events per participant. Table 1 lists the results of the Benjamini-Hochberg corrected paired t-tests for all six BIP events for both HiFi and LoFi conditions. All BIP events in both conditions showed a significant increase of EDA values between the investigated time frames. We analyzed HR and HRV derived from the BVP signal. To detect the effect on the HR after an event, we took a window of 3 s after the event and compared it with a 3 s baseline window before the event. For the OUTVRQ event we also compared both events (notification and HMD-off event) separately. In both fidelity conditions, neither the first blackout event nor the second BIP event showed a significant effect. Thus, the HR is not a reliable predictor for BIP events and we do not take the HR not into account in the further analysis of the BIP events. The same holds for the HRV. For the RR, we averaged and compared 10 s before and after the event as suggested by Liebold et al. [54]. However, the (Benjamini-Hochberg corrected) paired t-tests showed no significant differences for both blackout events in both fidelities.

Table 1 lists the results of the BIP investigations for EDA, HR, HRV and RR for every BIP event with the mean differences between the tested time frame (dependent on the signal) after the event and the baseline window before the event as well as the results of the corresponding paired samples t-test. The p-values for the t-test of each signal are Benjamini-Hochberg corrected. In line with [54] and [86], the detection of the BIP events was robust for the EDA signals for both fidelities. Thus, we base the further analysis on the EDA.

Further, we investigated which of the questionnaire modalities lead to the highest BIP as indicated by the strongest EDA response. We analyzed the mean difference between the two time frames before and after the BIP events, as used for BIP detection (c.f. Figure 4 and Table 1). Figure 5 shows barplots of the results. For both OUTVRQs we took the maximum of the response to the notification and HMD off event to gain a

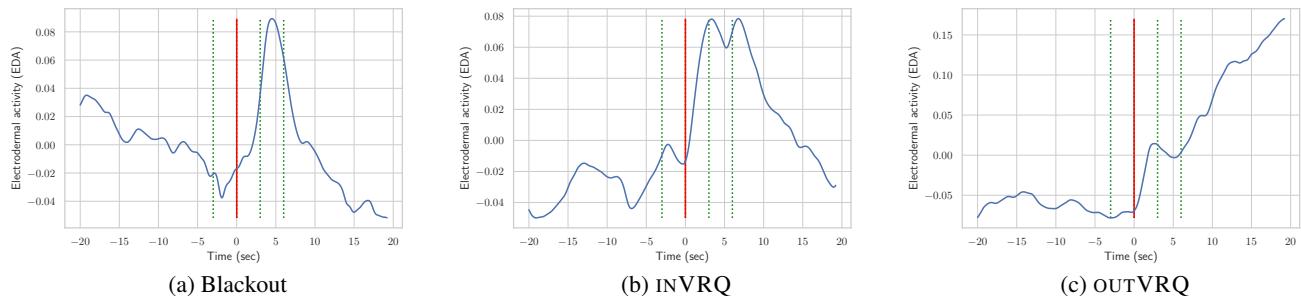


Figure 4. Normalized EDA reactions to all BIPs events. The x-axis represents time relative to the BIP event (red vertical line). The green dotted lines illustrate the windows used for event detection and the black line the “HMD off” event for the OUTVRQ. Note: Plot 4c has different scale-windows.

| | df | Blackout 1 | Blackout 2 | inVRQ 1 | inVRQ 2 | outVRQ 1 (notification) | outVRQ 1 (HMD off) | outVRQ 2 (notification) | outVRQ 2 (HMD off) |
|-------------|----|----------------|--------------|----------------|----------------|----------------------------|-----------------------|----------------------------|-----------------------|
| HiFi | | | | | | | | | |
| EDA | 23 | 0.05(2.23)* | 0.12(2.88)* | 0.04(2.21)* | 0.04(2.18)* | 0.05(2.28)* | 0.16(2.32)* | 0.09(2.57)* | 0.25(2.50)* |
| HR | 24 | -7.46(-2.38) | -4.09(-1.31) | -5.91(-2.38) | 0.64(0.23) | -2.65(-0.58) | -10.55(-2.15) | -1.78(-0.65) | 1.89(0.38) |
| HRV | 24 | -5.38(-1.68) | -1.16(-0.43) | -4.94(-1.72) | -1.63(-0.66) | -3.98(-1.22) | 5.65(0.96) | 1.85(0.89) | 9.00(1.56) |
| RR | 24 | -3.58(-3.02)** | -2.13(-1.93) | -2.51(-1.58) | -2.86(-1.76) | -0.12(-0.08) | -0.10(-0.07) | -1.84(-1.21) | -2.68(-1.75) |
| LoFi | | | | | | | | | |
| EDA | 24 | 0.06(2.49)* | 0.15(3.01)* | 0.15(3.08)* | 0.08(2.56)* | 0.11(2.12)* | 0.28(2.39)* | 0.07(1.60) | 0.20(2.81)* |
| HR | 24 | -2.23(-0.52) | 0.67(0.23) | 1.58(0.46) | -2.21(-0.67) | 1.69(0.63) | -1.88(-0.37) | 2.89(0.87) | 2.42(0.56) |
| HRV | 24 | -3.80(-1.46) | -2.27(-0.58) | -4.46(-1.44) | -1.43(-0.78) | 1.49(0.47) | 14.37(2.47)* | -2.21(-1.03) | 8.79(1.39) |
| RR | 24 | 0.55(0.40) | -2.27(-1.28) | -4.02(-4.10)** | -3.45(-4.01)** | -1.72(-1.31) | 2.56(1.94) | 1.07(0.99) | 3.52(3.36)* |

Table 1. Responses to every BIP event for EDA, HR, HRV, RR signals split by the fidelity condition. The OUTVRQ event is split into the occurrence of the notification and the HMD take off event. For each event the mean difference between the baseline window before the event and the test window after the event and the result of a paired-sampled t-test are given. P-values are corrected with Benjamini-Hochberg correction for each signal (row). For each value: Mean difference (t-statistic), ** p<0.01, * p<0.05

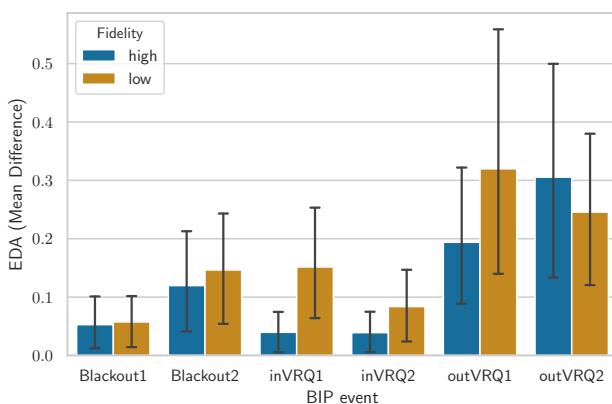


Figure 5. EDA differences (M,SD) of all BIP events for both fidelities.

single value. Anderson-Darling and a Levene tests assured that normality and equal variance assumptions were not violated ($\alpha=.05$) for all BIP events. To compare INVRQ and OUTVRQ reactions, we performed a mixed factorial ANOVA with the order of the questionnaire assessment (Q#1, Q#2) and the questionnaire modality (INVRQ, OUTVRQ) as within-subject factors and fidelity (HiFi, LoFi) as between-subject factor. The outcomes reveal a significant difference for questionnaire modality ($F(1,47)=11.71, p=.002, \eta_p^2=.33$, observed power $1-\beta=.99$) and an interaction effect between the questionnaire order and the fidelity ($F(1,47)=4.53, p=.04, \eta_p^2=.17$,

observed power $1-\beta=.89$). Fidelity did not show a main effect and we did not observe additional interactions. For questionnaire modality and questionnaire order*fidelity we performed post-hoc tests with Benjamini-Hochberg correction. For order*fidelity none of the post-hoc tests led to a significant difference. For the questionnaire modality the post-hoc test revealed that the reaction to the OUTVRQs led to a higher BIP than the INVRQ events ($t(97)=-4.14, p<.001$, Cohen's $d=.58$). Further we compared the reaction to the blackout events with the reaction to the INVRQ events, running a RM-ANOVA with the four INVRQ and blackout events as within-subject factor and fidelity as a between-subject factor. Mauchly's test indicated that the assumption of sphericity is violated ($\chi^2(5)=15.16, p=.01$), therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\epsilon=.83$). There was no significant difference between OUTVRQs and blackout events ($F(2.48, 116.64)=2.74, p=.06, \eta_p^2=.055, 1-\beta=.61$). The between factor fidelity ($F(1,47)=2.33, p=.13, \eta_p^2=.047, 1-\beta=.53$) and interaction ($F(2.48, 116.63)=1.14, p=.33, \eta_p^2=.024, 1-\beta=.28$) also did not reveal an effect.

Further, we examined the EDA recovery after BIP events. To this end, we applied a 3 s sliding window for a timeframe up to 50 s after the BIP until the signal sunk back to baseline. For the blackout events and for INVRQ2, the signal stabilized after 3 to 6 seconds. The recovery after INVRQ1 took longer. We observed significant differences to baseline for 3 – 9 s.

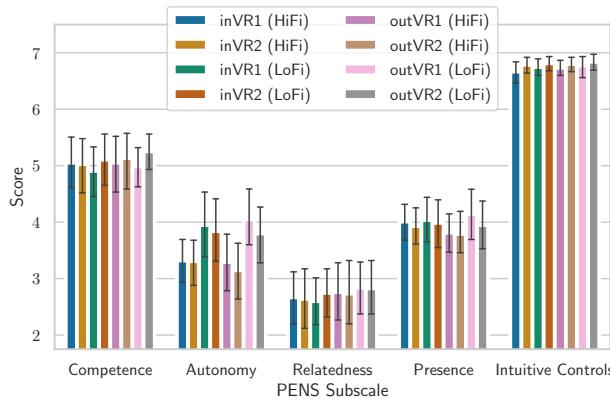


Figure 6. Results of the PENS questionnaires (M,SD).

Recovery from the OUTVRQs took by far the longest. Except for two windows of 6–9 s in LoFi and 9–12 s in HiFi (which we consider as outliers) the EDA signal was significantly above the baseline window for over 45 s. The effect can also be observed in the mean signal (cf. Fig. 4c).

Physiological evaluation of tasks

To investigate how the question asking method influences the subsequent tasks, we compared the average physiological signals during a task directly after the assessment of the subjective measurement for the previous task using either INVRQs or OUTVRQs. We conducted a two-way ANOVA with questionnaire modality and fidelity as factors to investigate the differences in physiological responses dependent on the preceded questionnaire modality. We found a main effect on questionnaire modality $F(1, 191)=28.30, p<.001, \eta^2=.99, 1-\beta=.99$ but no main effect on fidelity and no interactions. The post-hoc test confirmed this difference for tasks after an OUTVRQ ($M=.40, SD=.63$) and an INVRQ ($M=.004, SD=.41$) with $t(193)=-5.35, p<.001$ with a medium effect size of Cohen's $d=.77$. The other biosignals did not lead to consistent effects. A comparison of the biosignals during the actual filling out of INVRQs and OUTVRQs was not possible, as the activity of both tasks is structurally and behaviorally too different [49].

Subjective measurements

Figure 6 shows the results of all PENS takes for both questionnaires (INVRQ and OUTVRQ) and both fidelities (HiFi and LoFi). One sample t-tests against a neutral score of 4 revealed positive ratings on competence ($t(199)=13.25, p<.001$) and intuitive control ($t(199)=101.88, p<.001$), and negative scores on autonomy ($t(199)=-4.43, p<.001$) and relatedness ($t(199)=-14.40, p<.001$) within both fidelities. For all PENS subscales we conducted a mixed factorial ANOVA with the order of the repeated questionnaire (Q#1, Q#2) and the questionnaire modality (INVRQ, OUTVRQ) as within-subject factors and fidelity (HiFi, LoFi) as between-subject factor. The analysis revealed significant differences between Q#1 and Q#2 for autonomy ($F(1, 48)=5.89, p=.02, \eta^2=.11$, observed power $1-\beta=.68$), competence ($F(1, 48)=4.41, p=.04, \eta^2=.08, 1-\beta=.56$) and intuitive control ($F(1, 48)=8.90, p=.004$,

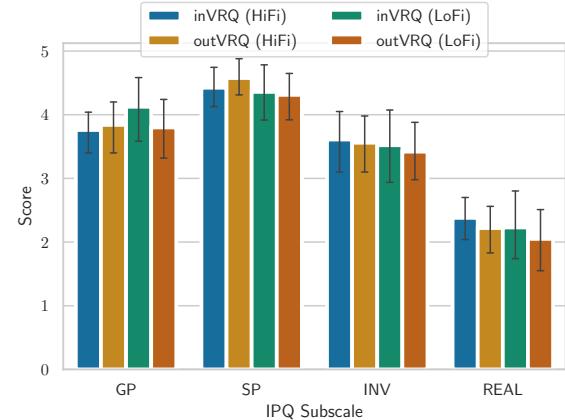


Figure 7. Post-experience IPQ results (M,SD). The REAL scores are below neutral for both fidelities, while all other scores are positive.

| | Cronbach's α outVR | Cronbach's α inVR | diff. | p |
|--------------------|------------------------------|-----------------------------|--------|------|
| Competence | 0.88 | 0.88 | 0.0035 | 0.45 |
| Autonomy | 0.85 | 0.85 | 0.0018 | 0.48 |
| Relatedness | 0.69 | 0.70 | 0.0161 | 0.41 |
| Presence | 0.84 | 0.84 | 0.0041 | 0.45 |
| Intuitive Controls | 0.75 | 0.73 | 0.0226 | 0.36 |

Table 2. Results of the Feldt reliability comparison [27, 28] between INVRQs and OUTVRQs for the PENS subscales.

$\eta_p^2=.15, 1-\beta=.85$). However, we did not find any significant effects/interactions on questionnaire modality or fidelity. Bonferroni-Holm corrected post-hoc t-tests between OUTVRQ1 and OUTVRQ2 showed a significant decrease of autonomy ($t(49)=2.33, p=.02$, Cohen's $d=.34$) and significant increases for intuitive control ($t(49)=-2.33, p=.002$, Cohen's $d=-.18$). As suggested by the ANOVA, a trend was indicated for competence ($t(48)=-1.95, p=.06$). To investigate random biases of the assessment method (H_4), we calculated the Cronbach's alphas of the PENS measures for both questionnaire modalities (Table 2). We compared the Cronbach's alphas using Feldt's test [27, 28] which showed no significant differences between INVRQs and OUTVRQs on any subscale.

For the post-experience IPQs, we conducted a mixed-factorial ANOVA with questionnaire modality as withing factor and fidelity as between factor. The analysis showed no significant main effects or interaction effects. However, one sample t-tests against neutral score (3) revealed positive ratings of GP ($t(99)=8.16, p<.001$), SP ($t(99)=15.59, p<.001$), and INV ($t(99)=4.13, p<.001$) and a negative score on REAL ($t(99)=-6.96, p<.001$) for both fidelities. The results are depicted in Figure 7.

Times & Performance

The completion time of the game experience including the time for filling out the PENS questionnaires was on average 742.78 s ($SD=78.31$) for the OUTVRQ condition and 680.90 s ($SD=87.00$) for the INVRQ condition with a significant difference ($t(99)=3.71, p < .001$). To exam-

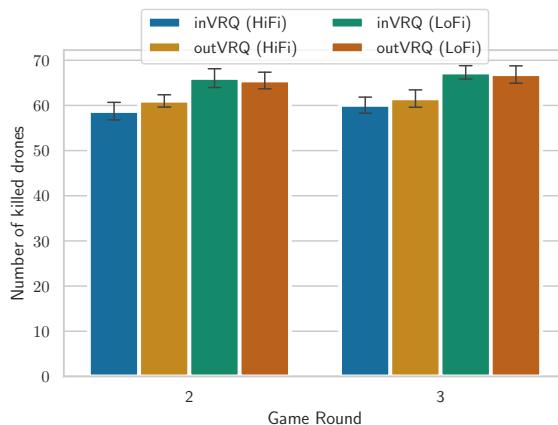


Figure 8. The performance (M,SD) for game rounds 2 and 3. For the HiFi condition, the scores are significantly lower in both rounds.

In the influence of the questionnaire modality on the performance in the subsequent task, we analyzed the number of eliminated drones for game rounds 2 and 3 (Figure 8). Anderson-Darling and Levene tests assured that normality and equal variance assumptions were not violated for the performance metrics ($\alpha=.05$). A mixed-factorial ANOVA with fidelity as a between-subject variable and the within-subjects variables questionnaire modality and scores in game round 2 and 3 showed no main effect for the questionnaire modality, but for fidelity ($F(1,48)=29.22, p<.001$, $\eta_p^2=.38, 1-\beta=.99$) and game round ($F(1,48)=6.14, p=.017$, $\eta_p^2=.11, 1-\beta=.70$). Post-hoc tests between HiFi and LoFi ($t(99)=-10.13, p<.001$, Cohen's $d=.20$), and game round 2 and 3 ($t(99)=-3.02, p=.003$, Cohen's $d=1.28$) confirm this difference. Additionally, we found an interaction effect for questionnaire modality*fidelity ($F(1,48)=5.39, p=.02$, $\eta_p^2=.10, 1-\beta=.64$). Yet, Benjamini-Hochberg corrected post-hoc tests, revealed no differences on questionnaire modality for HiFi ($t(98)=-2.01, p=.06$), nor for LoFi ($t(98)=.48, p=.63$).

Concluding questionnaire

At the end of each session, the participants ranked the interruptions during gameplay by their level of disturbance beginning with the strongest: In the LoFi condition, the users rated the blackout event as most disturbing factor ($M=2.04, SD=1.25$), second “Leaving the VR environment” ($M=2.36, SD=1.44$) followed by “Answering questionnaires in general” ($M=2.96, SD=1.25$), “The recording of the biosignals” ($M=4.12, SD=1.14$) and “Noise from outside” ($M=4.20, SD=.98$) with no significant difference between the three investigated BIPs (blackout, questionnaires, leaving VR). Thus, the most interrupting factors were the planned interruptions and not the recording setup or external noise. In the HiFi condition, the general order of the average ranking was the same, however a paired samples t-test revealed that leaving the VE was ranked significantly higher than answering questionnaires in general ($t(49)=4.19, p < .01$).

DISCUSSION

An essential precondition for evaluating the effects of breaking a virtual experience is to provide the participants with a virtual

experience which can be broken. The IPQ scores are high on General Presence, Spatial Presence and Involvement, but surprisingly below neutral on the Realness subscale. Remarkably, as in the results by Schwind et al. [74], the IPQ ratings were unaffected by the fidelity. The PENS outcomes revealed high scores on competence and intuitive control as it was intended in our design. Autonomy and relatedness were rated with lower scores, which was expected, since the game offered only one simple mechanic and no interaction with other players or virtual characters.

Neither game experience, nor prior experience in VR affected the ratings on IPQ and PENS. This asserts that the random assignment was successful. Further, the participants perceived the planned interruptions as much more disturbing than the uncontrolled factors, which indicates a valid study flow. Plausible ratings on the IPQ and PENS confirm that our design of the environment and the study are valid. Overall, we can conclude that the simple shooting game fulfilled its purpose and generated an engaging VR experience.

Our analysis of the physiological data (cf. Table 1 and Figure 4) show that the blackout events during gameplay and both questionnaire modalities induced significant BIPs expressed by phasic skin responses. With this, we confirm H₁ and support evidence of previous work [54, 58, 85, 86] that EDA is an effective predictor for BIPs. Countering intuition, HR failed to respond reliably to the blackout event, which is in line with related work on physiological BIP detection [58], but in contrast to other results [54, 85, 86]. We explain this missing effect with the physically demanding task: To protect all of the three crystals on the plateau and kill as many drones as possible, the participants had to turn around frequently. All BIP events lead to stopping of movement, i.e. the movement effect on the signal should not differ between conditions and thus not influence any comparisons. The measured drop in HR after stopping movement was higher than the expected amplitude after the BIP [54]. Thus, it negates a BIP effect on HR. Stopping movement influences all signals including the EDA, but not equally. So the expected effect of stopped movement on the EDA signal is a drop in amplitude [72]. Instead, we observe an increase in amplitude which is smaller than the BIP-triggered increase reported by Liebold et al. [54] (stationary). Thus, we conclude that the effect of BIP is dominant and counteracts the effect of stopping movement.

Regarding H₂, the comparisons of the BIPs induced by filling out questionnaires shows evidence that OUTVRQs elicit a stronger BIP than INVRQs. As depicted in Figure 5, OUT-VRQs produced higher EDA responses than INVRQs with an effect size of Cohen's $d=.77$. Although this is only a medium effect size, we measured it in a realistic setup not forcing for exaggerated BIPs. These differences might become more prominent as participants are immersed in longer VR sessions or with more demanding tasks.

The EDA response to the OUTVRQ events differ significantly from the responses to the blackouts and INVRQ events. Moreover, the ranking of disturbances indicate that the higher visual fidelity gives room for a stronger break in presence, while breaks in VE with a weaker sense of presence are less promi-

ment. In terms of temporal dynamics, the BIP appears more sudden in the blackout and INVRQ. While the EDA signals after INVRQs show similar patterns to responses after the blackouts and pacified within a few seconds, OUTVRQs stayed significantly above baseline for over 55 s (Fig. 4). These outcomes are in line with observations on the experience of exiting VR by Knibbe et al. [48].

There was no measurable difference on performance (number of killed drones) in successive game rounds. Thus, we cannot find direct evidence to support H₃. However, we measured a higher and longer sustained EDA for OUTVRQs. This might affect the performance for more complex or demanding tasks.

We could not find conclusive evidence to support H₄. The differences only between OUTVRQ1 and OUTVRQ2 on PENS indicate that OUTVRQs appear to be more sensitive to BIPs than INVRQ. This is also supported by a raised EDA in a successive game round after an OUTVRQ. However, the analysis of the reliability of PENS does not further substantiate these results. For both modalities the reliability was high and did not differ significantly.

LIMITATIONS AND FUTURE WORK

Most of our participants were students and only 16 % self-identified as female. Although we payed attention to distribute self-identifying females equally across the conditions, this indicates room for further validation. Our study design aimed to resemble a representative VR user study; therefore we asked our participants to sit at the PC when filling out questionnaires. In VR however, the participants were standing and moving around which is common in many VR applications and games [51] and was kept as such to retain ecological validity. Thus, we did not compare the physiological responses during filling out phase between the INVRQs and OUTVRQs. Such an investigation would be beneficial to examine the influence of the questionnaire modality on the attention or concentration during filling out the questionnaire.

For this study, we employed a shooting game which only consisted of a simple mechanic. Therefore, the setup might not be sensitive to effects of BIP on performance. Future work should investigate how BIPs affect performance of tasks with different degrees of complexity and workload. Likewise, we investigated the effects of BIP with only one specific application. Although the VE (particularly LoFi) was designed generically, it is unclear how presence can be affected in other types of VEs (e.g. learning, productivity, therapy) and should be investigated in the future. Furthermore, the participants interacted only with questionnaires containing Likert scales. Other types of questions (e.g. open-ended, single or multiple choice), other questionnaire types or extents might reveal different results. Unexpectedly, despite only altering the visuals and the sounds between LoFi and HiFi while keeping the game mechanics consistent, subjects in LoFi performed significantly better and hit more targets. A possible explanation is that missing aural feedback and less visual richness in LoFi, allowed for stronger focus on the task.

Our results show that INVRQs can reduce BIP, but not eliminate it entirely. Due to the design and length of the embedded

questionnaire this is not surprising, since the participants perceived the answering of the questionnaires as interrupting. In future work we will examine different contextualisation methods of INVRQs that foster a stronger connection between the questionnaires and the VR tasks and therefore, might reduce the question-asking related BIPs even further.

CONCLUSION

Questionnaires are widespread measurement instruments to assess subjective responses on a particular experience in VR user studies. However, research on presence shows evidence that switching between VR and physical reality leads to a break in presence [48, 89] that might alter the outcomes of the self-reports [74], especially when assessing constructs that are sensitive to disturbance and should be assessed with the least possible invasion. Researchers started to administer questionnaires in VR, which most likely reduce the BIP [29]. Yet, related work offered no clear evidence whether INVRQs diminish the BIP and to what extent filling out questionnaires contributes to a BIP.

In this paper, we investigated how question-asking itself breaks the VR experience (H₁) and if INVRQs can minimize the break in presence (H₂) and therefore, reduce uncontrolled biases in performance (H₃) and self-reports (H₄). To investigate our hypotheses, we conducted a user study ($n=50$) where we recorded biosignals of subjects while they played a simple VR shooter either with low or high visual fidelity and filled out questionnaires regarding their PX both in VR and on PC in physical reality. Our results clearly show that both questionnaire modalities produce BIPs. Moreover, the physiological responses in the EDA for INVRQs are significantly lower and shorter than to OUTVRQs. These results show evidence that INVRQs are less invasive than OUTVRQs and provide more reliable self-reports. Our findings suggest an influence of BIPs on performance of subsequent tasks. Moreover, these effects might become even more prominent in VEs high-quality VR experiences, such as AAA games as there is “more immersion to break”. Our findings can help researchers and designers to apply the appropriate instruments for their study design and lay groundwork for the design of INVRQs to provide validated and standardized methods of question-asking in VR.

ACKNOWLEDGMENTS

This research was partially funded by the European Unions’ Horizon 2020 research and innovation program (No 688244), project first.stage, and by the German Research Foundation DFG through the project “PALMS” (no. 402779631), as well as as part of Collaborative Research Center (Sonderforschungsbereich) 1320 “EASE – Everyday Activity Science and Engineering”, University of Bremen (<https://ease-crc.org/>). The research was conducted in subproject P01 *Embodied semantics for the language of action and change*.

REFERENCES

- [1] Elinda Ai-Lim Lee, Kok Wai Wong, and Chun Che Fung. 2010. How Does Desktop Virtual Reality Enhance Learning Outcomes? A Structural Equation Modeling Approach. *Computers & Education* 55, 4

- (Dec. 2010), 1424–1442. DOI:
<http://dx.doi.org/10/cgt64s>
- [2] Majed Al Zayer, Isayas B. Adhanom, Paul MacNeilage, and Eelke Folmer. 2019. The Effect of Field-of-View Restriction on Sex Bias in VR Sickness and Spatial Navigation Performance. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, 354:1–354:12. DOI:<http://dx.doi.org/10/gf3krr>
- [3] Dmitry Alexandrovsky, Susanne Putze, Michael Bonfert, Sebastian Höffner, Pitt Michelmann, Dirk Wenig, Rainer Malaka, and Jan D. Smeddinck. 2020. Examining Design Choices of Questionnaires in VR User Studies. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Honolulu, HI, USA. DOI:
<http://dx.doi.org/10.1145/3313831.3376260>
- [4] R.M. Baños, C. Botella, M. Alcañiz, V. Llano, 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:<http://dx.doi.org/10/fbg399>
- [5] Thomas Baumgartner, Dominique Speck, Denise Wettstein, Ornella Masnari, Gian Beeli, and Lutz Jäncke. 2008. Feeling Present in Arousing Virtual Reality Worlds: Prefrontal Brain Regions Differentially Orchestrate Presence Experience in Adults and Children. *Frontiers in Human Neuroscience* 2 (2008). DOI:<http://dx.doi.org/10/bq8w3p>
- [6] Mathias Benedek and Christian Kaernbach. 2010. A Continuous Measure of Phasic Electrodermal Activity. *Journal of Neuroscience Methods* 190, 1-5 (June 2010), 80–91. DOI:<http://dx.doi.org/10/cwn8gm>
- [7] Stéphane Bouchard, Stéphanie Dumoulin, Jeanne Talbot, André-Anne Ledoux, Jennifer Phillips, Johana Monthuy-Blanc, Geneviève Labonté-Chartrand, Geneviève Robillard, Matteo Cantamesse, and Patrice Renaud. 2012. Manipulating Subjective Realism and Its Impact on Presence: Preliminary Results on Feasibility and Neuroanatomical Correlates. *Interacting with Computers* 24, 4 (July 2012), 227–236. DOI:<http://dx.doi.org/10/f4c5pk>
- [8] Stéphane Bouchard, Julie St-Jacques, Geneviève Robillard, and Patrice Renaud. 2008. Anxiety Increases the Feeling of Presence in Virtual Reality. *Presence: Teleoperators and Virtual Environments* 17, 4 (July 2008), 376–391. DOI:<http://dx.doi.org/10/bctvw8>
- [9] Doug A. Bowman and Ryan P. McMahan. 2007. Virtual Reality: How Much Immersion Is Enough? *Computer* 40, 7 (July 2007), 36–43. DOI:
<http://dx.doi.org/10.1109/MC.2007.257>
- [10] Andrea Brogni, Vinoba Vinayagamoorthy, Anthony Steed, and Mel Slater. 2006. Variations in Physiological Responses of Participants During Different Stages of an Immersive Virtual Environment Experiment. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST '06)*. ACM, New York, NY, USA, 376–382. DOI:
<http://dx.doi.org/10/b9drht>
- [11] Andrea Brogni, Vinoba Vinayagamoorthy, Anthony Steed, and Mel Slater. 2007. Responses of Participants during an Immersive Virtual Environment Experience. *IJVR* 6, 2 (2007), 1–10.
- [12] Emily Brown and Paul Cairns. 2004. A Grounded Investigation of Game Immersion. In *Extended Abstracts of the 2004 Conference on Human Factors and Computing Systems - CHI '04*. ACM Press, Vienna, Austria, 1297. DOI:
<http://dx.doi.org/10.1145/985921.986048>
- [13] Karl-Erik Bystrom, Woodrow Barfield, and Claudia Hendrix. 1999. A Conceptual Model of the Sense of Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments* 8, 2 (1999), 241–244. DOI:<http://dx.doi.org/10/cn88p2>
- [14] Zekun Cao, Jason Jerald, and Regis Kopper. 2018. Visually-Induced Motion Sickness Reduction via Static and Dynamic Rest Frames. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 105–112. DOI:<http://dx.doi.org/10/gf39qh>
- [15] Albert S. Carlin, Hunter G. Hoffman, and Suzanne Weghorst. 1997. Virtual Reality and Tactile Augmentation in the Treatment of Spider Phobia: A Case Report. *Behaviour Research and Therapy* 35, 2 (Feb. 1997), 153–158. DOI:
<http://dx.doi.org/10/d3xr8m>
- [16] Alan Cheng, Lei Yang, and Erik Andersen. 2017. Teaching Language and Culture with a Virtual Reality Game. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 541–549. DOI:<http://dx.doi.org/10/gf78wz>
- [17] Tanvir Irfan Chowdhury, Sharif Mohammad Shahnewaz Ferdous, and John Quarles. 2017. Information Recall in a Virtual Reality Disability Simulation. ACM Press, 1–10. DOI:
<http://dx.doi.org/10.1145/3139131.3139143>
- [18] Mihaly Csikszentmihalyi. 1990. *Flow: The Psychology of Optimal Experience*. Harper & Row, New York.
- [19] James J. Cummings and Jeremy N. Bailenson. 2016. How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence. *Media Psychology* 19, 2 (April 2016), 272–309. DOI:
<http://dx.doi.org/10/gd3zd4>
- [20] Cath Dillon, Edmund Keogh, and Jonathan Freeman. 2002. It's Been Emotional: Affect, Physiology, and Presence. In *Proceedings of the Fifth Annual International Workshop on Presence, Porto, Portugal*.
- [21] Cath Dillon, Edmund Keogh, Jonathan Freeman, and Jules Davidoff. 2000. Aroused and Immersed: The Psychophysiology of Presence. In *Proceedings of 3rd*

- International Workshop on Presence, Delft University of Technology, Delft, the Netherlands.* 27–28.
- [22] Tilman Dingler, Kai Kunze, and Benjamin Outram. 2018. VR Reading UIs: Assessing Text Parameters for Reading in VR. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (CHI EA '18)*. ACM, New York, NY, USA, LBW094:1–LBW094:6. DOI: <http://dx.doi.org/10.1145/3170427.3188695>
- [23] H. Q. Dinh, N. Walker, L. F. Hodges, Chang Song, and A. Kobayashi. 1999. Evaluating the Importance of Multi-Sensory Input on Memory and the Sense of Presence in Virtual Environments. In *Proceedings IEEE Virtual Reality (Cat. No. 99CB36316)*. 222–228. DOI: <http://dx.doi.org/10/cstb55>
- [24] Alan Dix. 2009. Human-Computer Interaction. In *Encyclopedia of Database Systems*. Springer, 1327–1331.
- [25] John V. Draper, David B. Kaber, and John M. Usher. 1998. Telepresence. *Human Factors* 40, 3 (Sept. 1998), 354–375. DOI: <http://dx.doi.org/10/cnmgtw>
- [26] Stephen R. Ellis. 1995. Nature and Origins of Virtual Environments: A Bibliographical Essay. In *Readings in Human–Computer Interaction*, Ronald M. Baecker, Jonathan Grudin, William A. S. Buxton, and Saul Greenberg (Eds.). Morgan Kaufmann, 913–932. DOI: <http://dx.doi.org/10.1016/B978-0-08-051574-8.50096-0>
- [27] Leonard S. Feldt. 1969. A Test of the Hypothesis That Cronbach's Alpha or Kuder-Richardson Coefficient Twenty Is the Same for Two Tests. *Psychometrika* 34, 3 (Sept. 1969), 363–373. DOI: <http://dx.doi.org/10/d9br8k>
- [28] Leonard S. Feldt and Seonghoon Kim. 2006. Testing the Difference Between Two Alpha Coefficients With Small Samples of Subjects and Raters. *Educational and Psychological Measurement* 66, 4 (Aug. 2006), 589–600. DOI: <http://dx.doi.org/10/bj9rrf>
- [29] A. S. Fernandes and S. K. Feiner. 2016. Combating VR Sickness through Subtle Dynamic Field-of-View Modification. In *2016 IEEE Symposium on 3D User Interfaces (3DUI)*. 201–210. DOI: <http://dx.doi.org/10/gcmfn6>
- [30] Jonathan Freeman, S. E. Avons, Ray Meddis, Don E. Pearson, and Wijnand IJsselsteijn. 2000. Using Behavioral Realism to Estimate Presence: A Study of the Utility of Postural Responses to Motion Stimuli. *Presence: Teleoperators and Virtual Environments* 9, 2 (April 2000), 149–164. DOI: <http://dx.doi.org/10/drkr8>
- [31] Jonathan Freeman, S. E. Avons, Don E. Pearson, and Wijnand A. IJsselsteijn. 1999. Effects of Sensory Information and Prior Experience on Direct Subjective Ratings of Presence. *Presence: Teleoperators and*
- Virtual Environments* 8, 1 (Feb. 1999), 1–13. DOI: <http://dx.doi.org/10/dg7fn6>
- [32] Julian Frommel, Michael Weber, Katja Rogers, Julia Brich, Daniel Besserer, Leonard Bradatsch, Isabel Ortinau, Ramona Schabenberger, Valentin Riemer, and Claudia Schrader. 2015. Integrated Questionnaires: Maintaining Presence in Game Environments for Self-Reported Data Acquisition. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15*. ACM Press, London, United Kingdom, 359–368. DOI: <http://dx.doi.org/10/gfvjm6>
- [33] Google. 2018. Rendering Text in WebVR | Web. <https://developers.google.com/web/showcase/2017/within>. (Sept. 2018).
- [34] Google Developers. 2017. Designing Screen Interfaces for VR (Google I/O '17). (2017).
- [35] Nathan Navarro Griffin, James Liu, and Eelke Folmer. 2018. Evaluation of Handsbusy vs Handsfree Virtual Locomotion. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '18)*. ACM, New York, NY, USA, 211–219. DOI: <http://dx.doi.org/10/gf3ngm>
- [36] K. Gruchalla. 2004. Immersive Well-Path Editing: Investigating the Added Value of Immersion. In *IEEE Virtual Reality 2004*. 157–164. DOI: <http://dx.doi.org/10/fjw5zr>
- [37] Sandra G. Hart. 1986. *NASA Task Load Index (TLX)*. Volume 1.0; Paper and Pencil Package. Technical Report.
- [38] Carrie Heeter. 1992. Being There: The Subjective Experience of Presence. *Presence: Teleoperators and Virtual Environments* 1, 2 (1992), 262–271. DOI: <http://dx.doi.org/10/gdskgf>
- [39] Claudia Hendrix and Woodrow Barfield. 1996. Presence within Virtual Environments as a Function of Visual Display Parameters. *Presence: Teleoperators and Virtual Environments* 5, 3 (Jan. 1996), 274–289. DOI: <http://dx.doi.org/10/gf798x>
- [40] Claudia Mary Hendrix. 1994. *Exploratory Studies on the Sense of Presence in Virtual Environments as a Function of Visual and Auditory Display Parameters*. Ph.D. Dissertation.
- [41] Larry F. Hodges, Barbara O. Rothbaum, Rob Kooper, Dan Opdyke, Thomas Meyer, Johannes J. de Graaff, James S. Williford, and Max M. North. 1994. Presence as The Defining Factor in a VR Application: Virtual Reality Graded Exposure in the Treatment of Acrophobia. (1994).
- [42] Wijnand A. IJsselsteijn, Huib de Ridder, Jonathan Freeman, and Steve E. Avons. 2000. Presence: Concept, Determinants, and Measurement. In *Electronic Imaging*, Bernice E. Rogowitz and Thrasyvoulos N. Pappas (Eds.). San Jose, CA, 520–529. DOI: <http://dx.doi.org/10/fgk32n>

- [43] Lutz Jäncke. 2009. Virtual Reality and the Role of the Prefrontal Cortex in Adults and Children. *Frontiers in Neuroscience* 3, 1 (May 2009). DOI: <http://dx.doi.org/10/fc6r82>
- [44] Jason Jerald. 2016. *The VR Book: Human-Centered Design for Virtual Reality* (first edition ed.). Number 8 in ACM Books. acm, Association for Computing Machinery, New York. OCLC: 945367096.
- [45] HyeongYeop Kang, Geonsun Lee, Seongsu Kwon, Ohung Kwon, Seongpil Kim, and JungHyun Han. 2018. Flotation Simulation in a Cable-Driven Virtual Environment – A Study with Parasailing. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 632:1–632:11. DOI: <http://dx.doi.org/10.1145/3173574.3174206>
- [46] Pankaj Khanna, Insu Yu, Jesper Mortensen, and Mel Slater. 2006. Presence in Response to Dynamic Visual Realism: A Preliminary Report of an Experiment Study. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST '06)*. ACM, New York, NY, USA, 364–367. DOI: <http://dx.doi.org/10/bc4n4m>
- [47] Tina Kjær, Christoffer B. Lillelund, Mie Moth-Poulsen, Niels C. Nilsson, Rolf Nordahl, and Stefania Serafin. 2017. Can You Cut It?: An Exploration of the Effects of Editing in Cinematic Virtual Reality. ACM Press, 1–4. DOI: <http://dx.doi.org/10.1145/3139131.3139166>
- [48] Jarrod Knibbe, Jonas Schjerlund, Mathias Petraeus, and Kasper Hornbæk. 2018. The Dream Is Collapsing: The Experience of Exiting VR. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 483:1–483:13. DOI: <http://dx.doi.org/10.1145/3173574.3174057>
- [49] E. Kristal-Boneh, M. Raifel, P. Froom, and J. Ribak. 1995. Heart Rate Variability in Health and Disease. *Scandinavian Journal of Work, Environment & Health* 21, 2 (1995), 85–95. DOI: <http://dx.doi.org/10/d6v9vk>
- [50] Heinz W. Krohne, Boris Egloff, Carl-Walter Kohlmann, and Anja Tausch. 1996. Untersuchungen Mit Einer Deutschen Version Der "Positive and Negative Affect Schedule"(PANAS). *Diagnostica-Göttingen-* 42 (1996), 139–156.
- [51] Alexander Kunst. 2019. Kinds of Video Games for VR Headsets U.S. 2017. <https://www.statista.com/forecasts/790465/kinds-of-video-games-us-consumers-would-like-to-use-a-vr-headset-for>. (2019).
- [52] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. 2017. *Research Methods in Human-Computer Interaction* (2 ed.). Morgan Kaufmann, Cambridge, MA.
- [53] Jane Lessiter, Jonathan Freeman, Edmund Keogh, and Jules Davidoff. 2001. A Cross-Media Presence Questionnaire: The ITC-Sense of Presence Inventory. *Presence: Teleoperators and Virtual Environments* 10, 3 (June 2001), 282–297. DOI: <http://dx.doi.org/10/cgzd92>
- [54] Benny Liebold, Michael Brill, Daniel Pietschmann, Frank Schwab, and Peter Ohler. 2017. Continuous Measurement of Breaks in Presence: Psychophysiology and Orienting Responses. *Media Psychology* 20, 3 (July 2017), 477–501. DOI: <http://dx.doi.org/10/gf7c8f>
- [55] Matthew Lombard and Theresa Ditton. 1997. At the Heart of It All: The Concept of Presence. *Journal of Computer-Mediated Communication* 3, 2 (Sept. 1997). DOI: <http://dx.doi.org/10/dc7f6s>
- [56] Matthew Lombard, Theresa B Ditton, Daliza Crane, Bill Davis, Gisela Gil-Egui, Karl Horvath, and Jessica Rossman. 2000. Measuring Presence: A Literature-Based Approach to the Development of a Standardized Paper-and-Pencil Instrument. In *Third International Workshop on Presence, Delft*. Delft, Netherlands, 2–4.
- [57] Katerina Mania and Alan Chalmers. 2000. *A User-Centred Methodology for Investigating Presence and Task Performance*. Technical Report. University of Bristol, Bristol, UK, UK.
- [58] Michael Meehan. 2001. *Physiological Reaction as an Objective Measure of Presence in Virtual Environments*. Ph.D. Dissertation.
- [59] Michael Meehan, Brent Insko, Mary Whitton, and Frederick P. Brooks, Jr. 2002. Physiological Measures of Presence in Stressful Virtual Environments. In *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques*. ACM, New York, NY, USA, 645–652. DOI: <http://dx.doi.org/10/cqkjtk>
- [60] Marvin Minsky. 1980. Telepresence. (1980).
- [61] David Modjeska and John Waterworth. 2000. Effects of Desktop 3D World Design on User Navigation and Search Performance. In *2000 IEEE Conference on Information Visualization. An International Conference on Computer Visualization and Graphics*. 215–220. DOI: <http://dx.doi.org/10/fkwn4>
- [62] Michael Narayan, Leo Waugh, Xiaoyu Zhang, Pradyut Bafna, and Doug Bowman. 2005. Quantifying the Benefits of Immersion for Collaboration in Virtual Environments. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST '05)*. ACM, New York, NY, USA, 78–81. DOI: <http://dx.doi.org/10/bbfzfb>
- [63] Sebastian Oberdörfer, David Heidrich, and Marc Erich Latoschik. 2019. Usability of Gamified Knowledge Learning in VR and Desktop-3D. In *Proceedings of the*

- 2019 CHI Conference on Human Factors in Computing Systems (CHI '19).* ACM, New York, NY, USA, 175:1–175:13. DOI :<http://dx.doi.org/10/gf3k86>
- [64] Oculus. 2019. Introduction to Best Practices. <https://developer.oculus.com/design/latest/concepts/book-bp/>. (Sept. 2019).
- [65] Holger T. Regenbrecht, Thomas W. Schubert, and Frank Friedmann. 1998. Measuring the Sense of Presence and Its Relations to Fear of Heights in Virtual Environments. *International Journal of Human–Computer Interaction* 10, 3 (Sept. 1998), 233–249. DOI :<http://dx.doi.org/10/c2q35t>
- [66] Richard M. Ryan, C. Scott Rigby, and Andrew Przybylski. 2006. The Motivational Pull of Video Games: A Self-Determination Theory Approach. *Motivation and Emotion* 30, 4 (Nov. 2006), 344–360. DOI :<http://dx.doi.org/10/dpj8jp>
- [67] Maria V. Sanchez-Vives and Mel Slater. 2005. From Presence to Consciousness through Virtual Reality. *Nature Reviews Neuroscience* 6, 4 (April 2005), 332. DOI :<http://dx.doi.org/10/bh8pn6>
- [68] Corina Sas and Gregory M. P. O'Hare. 2003. Presence Equation: An Investigation into Cognitive Factors Underlying Presence. *Presence: Teleoperators and Virtual Environments* 12, 5 (2003), 523–537. DOI :<http://dx.doi.org/10/fc5zn5>
- [69] Klaus R. Scherer and Heiner Ellgring. 2007. Are Facial Expressions of Emotion Produced by Categorical Affect Programs or Dynamically Driven by Appraisal? *Emotion* 7, 1 (Feb. 2007), 113–130. DOI :<http://dx.doi.org/10/d6c9vn>
- [70] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. 2001. The Experience of Presence: Factor Analytic Insights. *Presence: Teleoper. Virtual Environ.* 10, 3 (June 2001), 266–281. DOI :<http://dx.doi.org/10/cvt2nz>
- [71] Peter Schulz, Dmitry Alexandrovsky, Felix Putze, Rainer Malaka, and Johannes Schöning. 2019. The Role of Physical Props in VR Climbing Environments. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, 183:1–183:13. DOI :<http://dx.doi.org/10/gf3kx3>
- [72] Johannes Schumm, Marc Bächlin, Cornelia Setz, Bert Arnrich, Daniel Roggen, and Gerhard Tröster. 2008. Effect of Movements on the Electrodermal Response after a Startle Event. *Methods of Information in Medicine* 47, 3 (2008), 186–191. DOI :<http://dx.doi.org/10.3414/ME9108>
- [73] Valentin Schwind, Pascal Knierim, Lewis Chuang, and Niels Henze. 2017. "Where's Pinky?": The Effects of a Reduced Number of Fingers in Virtual Reality. ACM Press, 507–515. DOI :<http://dx.doi.org/10.1145/3116595.3116596>
- [74] Valentin Schwind, Pascal Knierim, Nico Haas, and Niels Henze. 2019. Using Presence Questionnaires in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, 360:1–360:12. DOI :<http://dx.doi.org/10/gfz8sr>
- [75] Valentin Schwind, Pascal Knierim, Cagri Tasci, Patrick Franczak, Nico Haas, and Niels Henze. 2017. "These Are Not My Hands!": Effect of Gender on the Perception of Avatar Hands in Virtual Reality. ACM Press, 1577–1582. DOI :<http://dx.doi.org/10.1145/3025453.3025602>
- [76] Valentin Schwind, Sven Mayer, Alexandre Comeau-Vermeersch, Robin Schweigert, and Niels Henze. 2018. Up to the Finger Tip: The Effect of Avatars on Mid-Air Pointing Accuracy in Virtual Reality. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '18)*. ACM, New York, NY, USA, 477–488. DOI :<http://dx.doi.org/10/gf3nqj>
- [77] Fredric Shaffer, Didier Combatalade, Erik Peper, and Zachary M. Meehan. 2016. A Guide to Cleaner Electrodermal Activity Measurements. *Biofeedback* 44, 2 (June 2016), 90–100. DOI :<http://dx.doi.org/10/gf7c85>
- [78] Thomas B. Sheridan. 1992. Musings on Telepresence and Virtual Presence. *Presence: Teleoperators and Virtual Environments* 1, 1 (Jan. 1992), 120–126. DOI :<http://dx.doi.org/10/gdcftg>
- [79] Ben Shneiderman, Catherine Plaisant, Maxine Cohen, Steven Jacobs, Niklas Elmquist, and Nicholas Diakopoulos. 2016. *Designing the User Interface: Strategies for Effective Human-Computer Interaction* (6 edition ed.). Pearson, Boston.
- [80] Michael J Singer, Jennifer Ehrlich, and Stephen Cinq-Mars. 1995. *Task Performance in Virtual Environments: Stereoscopic versus Monoscopic Displays and Head-Coupling*. Technical Report. ARMY RESEARCH INST FOR THE BEHAVIORAL AND SOCIAL SCIENCES ALEXANDRIA VA.
- [81] Richard Skarbez, Frederick P. Brooks, Jr., and Mary C. Whitton. 2017. A Survey of Presence and Related Concepts. *ACM Comput. Surv.* 50, 6 (Nov. 2017), 96:1–96:39. DOI :<http://dx.doi.org/10/gfz8jf>
- [82] Mel Slater. 2003. A Note on Presence Terminology. *Presence connect* 3, 3 (2003), 1–5.
- [83] Mel Slater. 2004. How Colorful Was Your Day? Why Questionnaires Cannot Assess Presence in Virtual Environments. *Presence: Teleoperators & Virtual Environments* 13, 4 (2004), 484–493. DOI :<http://dx.doi.org/10/d7nqft>
- [84] Mel Slater. 2007. The Concept of Presence and Its Measurement. (July 2007).

- [85] Mel Slater, Andrea Brogni, and Anthony Steed. 2003. Physiological Responses to Breaks in Presence: A Pilot Study. In *Annual International Workshop On Presence*, Vol. 6. 4.
- [86] Mel Slater, Christoph Guger, Guenter Edlinger, Robert Leeb, Gert Pfurtscheller, Angus Antley, Maia Garau, Andrea Brogni, and Doron Friedman. 2006. Analysis of Physiological Responses to a Social Situation in an Immersive Virtual Environment. *Presence: Teleoperators and Virtual Environments* 15, 5 (Oct. 2006), 553–569. DOI :<http://dx.doi.org/10/bgcpj7>
- [87] Mel Slater, Pankaj Khanna, Jesper Mortensen, and Insu Yu. 2009. Visual Realism Enhances Realistic Response in an Immersive Virtual Environment. *IEEE Computer Graphics and Applications* 29, 3 (May 2009), 76–84. DOI :<http://dx.doi.org/10/d6vn2k>
- [88] Mel Slater, Vasilis Linakis, Martin Usoh, and Rob Kooper. 1996. Immersion, Presence and Performance in Virtual Environments: An Experiment with Tri-Dimensional Chess. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST '96)*. ACM, New York, NY, USA, 163–172. DOI :<http://dx.doi.org/10/gf735d>
- [89] Mel Slater and Anthony Steed. 2000. A Virtual Presence Counter. *Presence: Teleoperators and Virtual Environments* 9, 5 (2000), 413–434. DOI :<http://dx.doi.org/10/cd5kxg>
- [90] Mel Slater, Martin Usoh, and Yiorgos Chrysanthou. 1995. The Influence of Dynamic Shadows on Presence in Immersive Virtual Environments. In *Virtual Environments '95 (Eurographics)*, Martin Göbel (Ed.). Springer Vienna, 8–21.
- [91] Mel Slater, Martin Usoh, and Anthony Steed. 1994. Depth of Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments* 3, 2 (1994), 130–144. DOI :<http://dx.doi.org/10/gf8bfp>
- [92] Mel Slater and Sylvia Wilbur. 1997. A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments* 6, 6 (1997), 603–616. DOI :<http://dx.doi.org/10/gd3zdp>
- [93] Michael P. Snow and Robert C. Williges. 1998. Empirical Models Based on Free-Modulus Magnitude Estimation of Perceived Presence in Virtual Environments. *Human Factors* 40, 3 (1998), 386–402. DOI :<http://dx.doi.org/10/cqrrt7>
- [94] Maurício Sousa, Daniel Mendes, Soraia Paulo, Nuno Matela, Joaquim Jorge, and Daniel Simões Lopes. 2017. VRROOM: Virtual Reality for Radiologists in the Reading Room. ACM Press, 4057–4062. DOI :<http://dx.doi.org/10.1145/3025453.3025566>
- [95] John Towell and Elizabeth Towell. 1997. Presence in Text-Based Networked Virtual Environments or “MUDS”. *Presence: Teleoperators and Virtual Environments* 6, 5 (Oct. 1997), 590–595. DOI :<http://dx.doi.org/10/gf8bhf>
- [96] Martin Usoh, Ernest Catena, Sima Arman, and Mel Slater. 2000. Using Presence Questionnaires in Reality. *Presence: Teleoperators and Virtual Environments* 9, 5 (2000), 497–503. DOI :<http://dx.doi.org/10/bf3cmp>
- [97] Marieke van Dooren, J. J. G. (Gert-Jan) de Vries, and Joris H. Janssen. 2012. Emotional Sweating across the Body: Comparing 16 Different Skin Conductance Measurement Locations. *Physiology & Behavior* 106, 2 (May 2012), 298–304. DOI :<http://dx.doi.org/10/gd3vxf>
- [98] Robert B. Welch. 1999. How Can We Determine If the Sense of Presence Affects Task Performance? <http://dx.doi.org/10.1162/105474699566387> (1999). DOI :<http://dx.doi.org/10.1162/105474699566387>
- [99] Robert B. Welch, Theodore T. Blackmon, Andrew Liu, Barbara A. Mellers, and Lawrence W. Stark. 1996. The Effects of Pictorial Realism, Delay of Visual Feedback, and Observer Interactivity on the Subjective Sense of Presence. *Presence: Teleoperators and Virtual Environments* 5, 3 (1996), 263–273. DOI :<http://dx.doi.org/10/gf796n>
- [100] Brenda K Wiederhold, Dong P Jang, Mayumi Kaneda, Yair Lurie, Todd May, In Y Kim, and Sun I Kim. 2003. An Investigation into Physiological Responses in Virtual Environments: An Objective Measurement of Presence. In *Towards Cyberpsychology: Mind, Cognitions and Society in the Internet Age*. IOS Press, Amsterdam, 10.
- [101] C. Wienrich, K. Schindler, N. 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. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 207–214. DOI :<http://dx.doi.org/10/gf39qb>
- [102] Bob G. Witmer and Michael J. Singer. 1998. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments* 7, 3 (June 1998), 225–240. DOI :<http://dx.doi.org/10/dw9b5s>
- [103] Pavel Zahorik and Rick L. Jenison. 1998. Presence as Being-in-the-World. *Presence: Teleoperators and Virtual Environments* 7, 1 (Feb. 1998), 78–89. DOI :<http://dx.doi.org/10/fsg6d6>
- [104] P. Zimmons and A. Panter. 2003. The Influence of Rendering Quality on Presence and Task Performance in a Virtual Environment. In *Proceedings IEEE Virtual Reality*. 293–294. DOI :<http://dx.doi.org/10/bkjdq5>