Towards an objective measure of presence: examining startle reflexes in a commercial virtual reality game

SCHIRM, Johannes, TULLIUS, Gabriella and HABGOOD, Jacob <http://orcid.org/0000-0003-4531-0507>

Published version

SCHIRM, Johannes, TULLIUS, Gabriella and HABGOOD, Jacob (2019). Towards an objective measure of presence: examining startle reflexes in a commercial virtual reality game. In: CHI PLAY ’19. Extended abstracts of the annual symposium on computer-human interaction in play. New York, USA, ACM, 671-678.

Towards an Objective Measure of Presence: Examining Startle Reflexes in a Commercial Virtual Reality Game

Johannes Schirm
Reutlingen University
Reutlingen, Germany
research@johannes-schirm.de

Gabriela Tullius
Reutlingen University
Reutlingen, Germany

Jacob Habgood
Steel Minions Game Studio
Sheffield Hallam University
Sheffield, United Kingdom

CHI'20,, April 25–30, 2020, Honolulu, HI, USA
ACM 978-1-4503-6819-3/20/04.
https://doi.org/10.1145/3334480.XXXXXXX

Abstract
A large body of literature is concerned with models of presence—the sensory illusion of being part of a virtual scene—but there is still no general agreement on how to measure it objectively and reliably. For the presented study, we applied contemporary theory to measure presence in virtual reality. Thirty-seven participants explored an existing commercial game in order to complete a collection task. Two startle events were naturally embedded in the game progression to evoke physical reactions and head tracking data was collected in response to these events. Subjective presence was recorded using a post-study questionnaire and real-time assessments. Our novel implementation of behavioral measures lead to insights which could inform future presence research: We propose a measure in which startle reflexes are evoked through specific events in the virtual environment, and head tracking data is compared to the range and speed of baseline interactions.

Author Keywords
Presence; Virtual reality; Subjective and behavioral measures; Head movement; Startle events; Game

CCS Concepts
•**Human-centered computing** → **Virtual reality;** *User studies;*

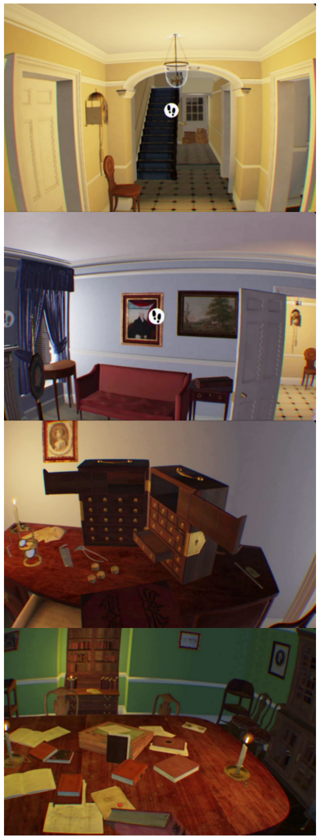

Figure 1: Four rooms from the commercial game *The Chantry* [19], which provided the basis for the presented study. A detailed reconstruction of a historical house in Georgian style was created for the original game.

Introduction

Presence is considered one of the most important factors contributing to user experience in virtual reality (VR), and its models and measurement are currently subject to intensive research. Still, obtaining an objective and reliable measure of presence is not easy. Existing methods either rely on subjective assessments by participants or on complex experimental setups involving non-standard hardware.

In their recent literature survey, Skarbez, Brooks and Whitton summarized the history of presence research and analyzed a large number of instruments. They recommend "the use [of] multiple measures of different types whenever feasible" and further argue that "behavioral measures represent a promising area of study that has so far been understudied" [12, p. 32]. Unlike physiological measures, they often integrate naturally into virtual environments, without disturbing users or making the procedure too complex. We observe that there is significantly less literature available on behavioral measures than there is on subjective or physiological measures, even though all of them have their own specific advantages and disadvantages.

Inspired by early ideas for reflex-based measures of presence, we think that analyzing head-tracking data could be an objective and easily implementable approach to this problem, especially in the context of games. Based on the most relevant work from this field, we designed an experiment which employs different measures of presence and offers participants an engaging, high-quality VR experience. With two existing measures of presence (subjective real-time assessments and a post-study questionnaire) for comparison, we evaluate our proposed startle reflex measure. This novel approach to measuring presence is mostly transparent to the user and requires nothing more than the transformation matrix of the VR headset.

Related Work

In our understanding, presence is the sensory illusion of "being part" of a virtual scene, which aligns with most views in the literature. Although this short definition of presence is rather straight-forward and can easily be rated by users in real-time and post-study, there were extensive discussions on how to define the concept more precisely.

On the most fundamental level, efforts were made to define presence through Gibson's model of affordance [4], through decomposition into immersion, presence and performance [2] and through social and perceptual aspects [6]. The most influential theoretical framework for presence by Slater and consists of *place illusion* and *plausibility* [14]. When it comes to actually measuring presence, there are three major approaches discussed and used within literature: Subjective, physiological and behavioral. In the following subsections, each of them is reviewed in more detail.

Subjective

The general challenge with subjective measures lies in the concept of presence used to design questions and in particular the way these are understood by participants. For example, Slater compared this to asking a person as how "colorful" they perceived yesterday [13, 15]. Still, there are many questionnaires available that have been shown to be "valid, sensitive, and reliable" [12, p. 28].

Most of these are post-study questionnaires, summarizing the participant's whole experience. Those by Slater, Usoh and Steed [16] and Witmer and Singer [23] are most commonly used, as well as the Igroup Presence Questionnaire [10]. Bouchard et al. propose a 1-item instrument [1], which is reported to be working reliably in spite of its simplicity: "Results show that the question is well-understood, reliable between tests for the same users, [...] and is sensitive between high and low levels of presence" [1, p. 27].

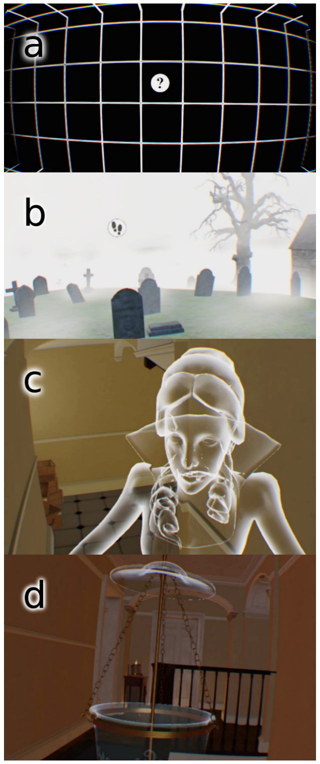

Figure 2: After a short briefing in a neutral environment (a), a short adaptation task on the graveyard in front of the house (b) had to be performed. During the main task, participants encountered two special events: A ghost in their private space (c) and a chandelier swinging towards their head (d). Both were expected to evoke stronger physical reactions in present participants.

Physiological

Even though physiological measures are most objective and provide real-time data, they often require special equipment attached to the participant, potentially influencing presence through distraction [12, p. 31]. They are most frequently used in virtual environments that "are known to affect physiological signals in certain ways" [12, p. 31], for example threatening or stressful situations. Meehan et al. [7, p. 650] found the difference in heart rate—the most distinct objective measure in their opinion—to correlate with the questionnaire by Slater, Usoh and Steed. Deniaud et al. used physiological measures when analyzing a driving simulation and concluded that "presence measurement can't be only based on subjective measures" [3].

Behavioral

Early ideas by Sheridan focused on natural reactions to specific cues, for example continuing to follow social conventions, even when experiencing a non-shared virtual environment, and dodging an object on a collision course [11, p. 4]. Nichols, Haldane and Wilson categorized observed reactions on startle events into physical, verbal and no reaction and found this to significantly correlate to the feeling of having "visited" the virtual environment [9, p. 478]. Only few behavioral measures like the *comeback rate* by Thie and van Wijk [20] are independent of the environment.

Experimental Design

Participants

This experiment was run in the context of the Horizon 2020 REVEAL project. 37 students (5 female, 32 male) from two technical courses at Sheffield Hallam University volunteered to take part in the experiment. Aged between 19 and 24 years (M=21.54, SD=1.28), they rated their experience on a scale of 0 to 6 to be 5.46 (SD=0.8) with video games and 3.08 (SD=1.83) with VR.

Virtual Reality Game

The Chantry is an educational Environmental Narrative [17, p. 34] game for PlayStation VR, developed by the *Steel Minions* game studio [19]. The virtual environment used in our study is a derivative of this game, the general framework for which has already been described by Habgood et al. [5, pp. 372–373]. Keeping the game's detailed environment (see Figure 1), its story was replaced by an experimental procedure which involves locating seven items. Environmental Narrative games typically provide a continuous, natural experience with a strong focus on presenting an immersive environment, ideal for measuring presence.

Participants navigated through the virtual environment on a predefined node-graph and were able to rotate their view in steps of 30 degrees using a DualShock 4 controller in their hands. Adjacent nodes were represented by a footstep icon in world-space. Users had to direct their head towards these nodes and press a button on the controller to start the transition. Node-based locomotion was chosen because Habgood et al. previously suggested that it combines the advantages of free locomotion (intuitive, but evokes motion sickness) and avatar-based teleportation (much less motion sickness, but also less intuitive) [5, p. 375].

Phases

Before participants immersed themselves in the detailed environment of the game, they were introduced to the experiment in a separate, neutral virtual room (see Figure 2a). After a short interaction test, participants were explained the concept of presence and how to give spoken assessments of their presence while playing. All instructions were given in the form of prerecorded audio. The first measurement was taken directly after the introduction. Due to the minimal nature of the preparation room, it was also used as the baseline for all following assessments.

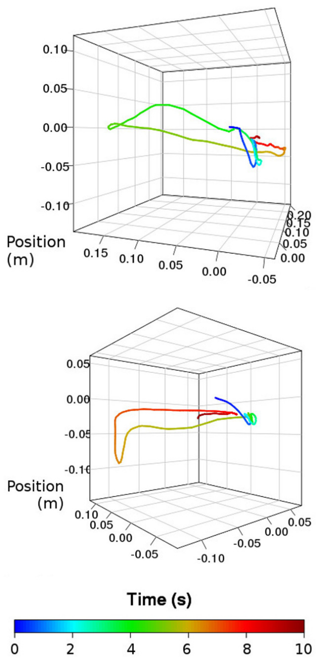

Figure 3: Two participants volunteered for a pilot experiment just with behavioral measures. The figure above visualizes their head movements while they encountered the ghost (see Figure 2c). The color-scale represents the elapsed time in seconds since their arrival at the locked door and the appearance of the ghost behind them. Positions are in meters.

To give participants time to adjust themselves to VR and the navigation system, a small task (M=3.48 minutes, SD=0.97) had to be completed on the graveyard in front of the house (see Figure 2b). Two baselines of normal head movement behavior were recorded at this opportunity. The game progression then led participants into the house. For a more natural experience, they could freely choose the order in which to collect the items. General game design principles were applied when developing this task to make a flow experience more likely. Finding all seven items took participants on average 11.04 minutes (SD=3.51).

Instruments

Out of available post-study questionnaires, the Presence Questionnaire (PQ) by Witmer and Singer [23] is one of the most established and is reported to work well with the real-time instrument by Bouchard et al. [1, p. 60]. The Immersive Tendencies Questionnaire (ITQ) by the same authors supplements the PQ with questions about personal conditions that could influence presence. Revised versions of the PQ (3.0) [22] and the ITQ (3.01) [21] from 2004 were used for this study. All items of the PQ and ITQ use a 7-point scale. Participants gave their answers through an electronic questionnaire application, one question at a time, using a slider ranging from 0 to 6 with one label for each extreme.

A post-study questionnaire cannot assess presence in real-time, so the 1-item instrument by Bouchard et al. [1] was also used for this study. At ten predefined points in the game, the prerecorded voice asked: “On a scale from zero to ten, to which extent do you feel present in the scene?” Using the HMD’s microphone, spoken assessments were given on the go, without pausing the game. It is criticized that repeated real-time assessments “intrude[...] on the very presence illusion one is trying to measure” [7], but this instrument is described as being comparatively unobtrusive.

Behavioral measures of presence are less common in the literature and thus much more rarely used. Similar to physiological measures, they are are difficult to interpret and require dedicated cues for evoking certain behavior to be integrated into the virtual environment. We refer to these cues as “special events.” Focusing on natural startle reflexes, we base our measure on rather low-level behavior. Authors such as Nichols et al. [9] explain why this could be a sensible way to indirectly determine presence.

We designed two special events: First, upon uncovering an item at the top of a chandelier, the mounting loosed and swung the chandelier (see Figure 2d) towards the user’s head. Second, upon turning back their view from a locked door at the end of a narrow hallway, users surprisingly faced a ghost (see Figure 2c) at just 70 cm of distance, certainly within their private space [18]. The ghost avoids the *Uncanney Valley* which a more realistic human model could fall into [8]. Both events are well within the requirements of the 12+ PEGI age rating of the original game, and were not scary after the initial startle effect. The headset’s transformation was recorded at 25 samples per second.

No standardized procedure is known to the authors that would be specifically capable of classifying the resulting trajectory, as the ideal approach would greatly depend on the specific event. As a first step, we aim for a visual analysis of recorded tracking data appropriate to each event. Figure 3 shows the reaction of two volunteers to the **ghost** special event during a pilot experiment. Both came from a technical background and were substantially involved in the development of the original game, yet they both reacted with a strong startle reflex, featuring a sudden acceleration and a temporary offset of 15-20 centimeters. However, unlike the other participants, they played the pilot experiment alone in the evening, with nobody else in the room.

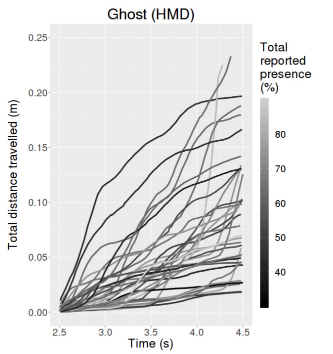

Figure 4: Accumulated head movement distance during the **chandelier** special event in meters. The color-scale codes real-time presence over the whole experiment in percent.

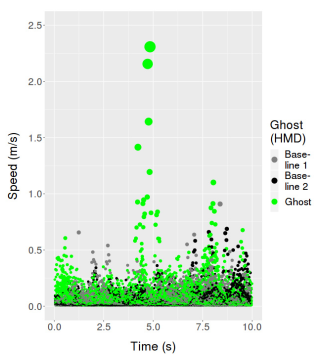

Figure 5: Head movement speeds during the **ghost** special event and the baseline data in meters per second. High values have bigger points for less overlap in low ones.

Procedure

Groups of 3-6 participants performed the experiment in parallel in a teaching lab, sitting on a chair, with a PSVR head-mounted display donned and a Dual Shock 4 controller in their hands. They had already signed an informed consent and answered general questions through an electronic questionnaire application, which they also used to answer the PQ directly after the experience ended.

Results

37 participants completed the post-questionnaire. 35 of them completed the main task without experiencing motion sickness and were considered in the analysis of tracking data and real-time assessments. Assessments were transcribed into text files for the 28 participants for which a complete set is available. It is possible that the others felt inhibited speaking out loud in the lab environment.

The total PQ score is 70.68 (SD=16.72) with a total ITQ score of 45.73 (SD=9). Examining the answer set with respect to self-evaluation of performance (e.g. question 15 and 16), participants reported that they adjusted to the virtual environment quickly (M=5.02; SD=0.98) and felt proficient (M=4.67; SD=1.151). This corroborates our view of the original VR game in terms of professionalism.

Real-time assessments were converted to δPresence using the first value given inside the neutral room (see Figure 2a). They are highly dispersed and show no apparent difference between measurements, but the δPresence means significantly correlate to their respective participant's PQ score ($r(26)=0.66$, $p=0.0001$, 95% confidence interval). 28.6 % of our participants reported very high presence (9 or 10), and 60 % reported high presence (8, 9 or 10) at least once. This indicates that the question was integrated in way that would not generally diminish presence in a significant way.

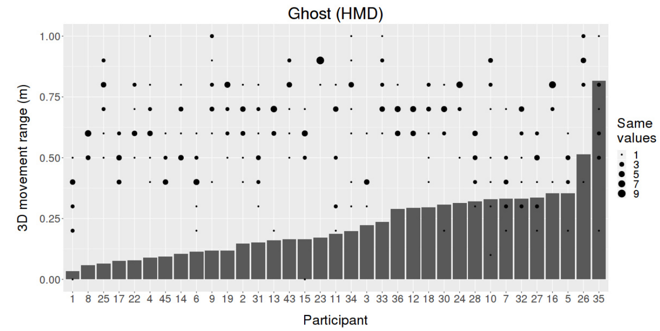

Figure 6: Dimensions (X+Y+Z) of the head movement bounding box during the *ghost* special event in meters. Available real-time presence assessments were normalized and overlaid.

The **chandelier** special event was expected to make participants instinctively move their head aside in order to prevent a "collision." This movement characteristic should be most meaningful when analyzed over time. Figure 4 visualizes for each participant how far their head-mounted display travelled during the time the chandelier swung towards them. For comparison, the normalized sum of available real-time assessments was color-coded for each participant.

The pilot experiment (see Figure 3) provides a well-founded starting point for the **ghost** special event. We expected a significant acceleration and an offset of about 20 cm. This movement characteristic should be most meaningful when analyzed in bounding box dimensions and speeds. Figure 6 shows the sum of all bounding box dimensions for each participant, considering 7.5 seconds of tracking data, and omitting about 2.5 seconds of orientation in front of the locked door. All reported real-time assessments were added per participant for comparison. Figure 5 visualizes movement speeds over time for all participants, with baseline speeds from the adaptation phase in the background.

	Range and speed	No reaction
Rating	**7.33** (3.08)	**6.76** (1.92)
Pres. (total)	**64.22** (15.11)	**60.8** (14.47)
Pres. (**ghost**)	**7.50** (1.58)	**6.74** (1.63)
Exp. (Games)	**5.33** (0.82)	**5.38** (0.86)
Exp. (VR)	**2.00** (1.55)	**2.86** (1.68)

Figure 7: A comparison of relevant meta data of participants whose reaction on the **ghost** special event was either classified as strong (reaction in range and speed of head movements) or none (reaction in neither range nor speed). Considered factors are a rating of the experiment scenario (11-point), real-time presence during the whole experiment and directly after the **ghost** special event and experience with video games and VR (7-point). Means are in bold face, standard deviations are below in brackets.

Discussion

Travelled distance is the only metric for the **chandelier** special event in our analysis because it accounts for all possible reactions where participants try to make room for the chandelier. With a variety of progression patterns, it is difficult to determine how strong reactions were just by looking at individual lines in Figure 4. Overall, stronger and weaker reaction patterns are equally present, and there seems to be no correlation to subjective presence.

For the **ghost** special event, speed and total range of movement are considered. Several participants featured an acceleration that can be clearly identified with the baseline as a reference. Similar to the total travelled distance metric of the **chandelier** special event, no justifiable baseline is available for the bounding box metric, making it difficult to put these observations into context. Here, too, there seems to be no correlation to absolute real-time presence values. However, at least for the presented visualizations, one can visually pinpoint reasonable reaction thresholds for movement range (bounding box) and movement speed. This can be done with more confidence using the exemplary reactions from the pilot experiment (see Figure 3) as a basis.

For the ranges in Figure 6, there is a noticeable jump between participants 33 and 36. For the speeds in Figure 5, green peaks represent unusual acceleration. With thresholds of 25 cm in range and 0.5 m per second in speed, reactions on the **ghost** special event are classified into the extremes “speed and range” and “no reaction.” Figure 7 shows a comparison of (for presence expectedly) relevant meta data such as video games and VR experience, a rating of the playing experience from 0 (poor) to 10 (excellent) and self-reported real-time presence in total and after facing the ghost. There is no significant difference in single factors, but there is a consistent overall tendency.

Participants whose reaction was classified as strong have more experience (mainly with VR), felt more present and were more impressed with our VR game. In summary, this supports our reaction thresholds. We conjecture that reasons for the ubiquitous disparity between subjective presence and behavior could be behavioral variance or even there being conscious and unconscious presence, the latter of which we would assume to cause startle responses.

Conclusion

The presented study evaluated feelings of presence in a high-quality VR game using a post-study questionnaire, real-time assessments and behavioral measures. By analyzing head movements, we explored a novel, unobtrusive and simple approach to measuring presence more objectively and reliably. Although our procedure of determining reaction thresholds would need validation in other contexts and we cannot discount the effect of other factors on this result (such as the research environment), we believe that our study provides valuable insights for future research.

Head movement speeds of a startle response produced the most distinct results in our study, especially in conjunction with bounding box dimensions of the head movement trajectory. Future research could generalize our special events, for example by testing whether the ghost could be replaced by a nonhuman object or how often startle responses can be provoked in the same scenario. It should also establish standardized procedures for determining reaction thresholds and recording reliable baselines.

Acknowledgements

Supported in part by the European Union’s Horizon 2020 research and innovation programme under grant agreement 732599. The authors would like to thank David Moore and Andrew Hamilton for their technical support.

REFERENCES

1. S. Bouchard, G. Robillard, J. St-Jacques, S. Dumoulin, M. J. Patry, and P. Renaud. 2004. Reliability and validity of a single-item measure of presence in VR. In *Proceedings. Second International Conference on Creating, Connecting and Collaborating through Computing*. 59–61. DOI: http://dx.doi.org/10.1109/HAVE.2004.1391882

2. 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/10.1162/105474699566107

3. C. Deniaud, V. Honnet, B. Jeanne, and D. Mestre. 2015. An investigation into physiological responses in driving simulators: An objective measurement of presence. In *Science and Information Conference (SAI)*. 739–748. DOI: http://dx.doi.org/10.1109/SAI.2015.7237225

4. John M. Flach and John G. Holden. 1998. The Reality of Experience: Gibson's Way. *Presence: Teleoperators and Virtual Environments* 7, 1 (1998), 90–95. DOI: http://dx.doi.org/10.1162/105474698565550

5. M. P. Jacob Habgood, David Moore, David Wilson, and Sergio Alapont. 2018. Rapid, Continuous Movement Between Nodes as an Accessible Virtual Reality Locomotion Technique. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 371–378. DOI: http://dx.doi.org/10.1109/VR.2018.8446130

6. Matthew Lombard and Theresa Ditton. 1997. At the Heart of It All: The Concept of Presence. *Journal of Computer-Mediated Communication* 3, 2 (1997). DOI: http://dx.doi.org/10.1111/j.1083-6101.1997.tb00072.x

7. Michael Meehan, Brent Insko, Mary Whitton, and Jr Brooks Frederick P. 2002. Physiological Measures of Presence in Stressful Virtual Environments. *ACM Transactions on Graphics* 21, 3 (July 2002), 645–652. DOI:http://dx.doi.org/10.1145/566654.566630

8. M. Mori, K. F. MacDorman, and N. Kageki. 2012. The Uncanny Valley [From the Field]. *IEEE Robotics Automation Magazine* 19, 2 (June 2012), 98–100. DOI: http://dx.doi.org/10.1109/MRA.2012.2192811

9. Sarah Nichols, Clovissa Haldane, and John R. Wilson. 2000. Measurement of presence and its consequences in virtual environments. *International Journal of Human-Computer Studies* 52, 3 (2000), 471. DOI: http://dx.doi.org/10.1006/ijhc.1999.0343

10. Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. 2001. The Experience of Presence: Factor Analytic Insights. *Presence: Teleoperators and Virtual Environments* 10, 3 (2001), 266–281. DOI: http://dx.doi.org/10.1162/105474601300343603

11. Thomas B. Sheridan. 1992. Musings on Telepresence and Virtual Presence. *Presence: Teleoperators and Virtual Environments* 1, 1 (1992), 120–126. DOI: http://dx.doi.org/10.1162/pres.1992.1.1.120

12. Richard Skarbez, Frederick P. Brooks, and Mary C. Whitton. 2017. A Survey of Presence and Related Concepts. *Comput. Surveys* 50, 6 (November 2017), 1–39. DOI:http://dx.doi.org/10.1145/3134301

13. Mel Slater. 2004. How Colorful Was Your Day? Why Questionnaires Cannot Assess Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments* 13, 4 (2004), 484–493. DOI: http://dx.doi.org/10.1162/1054746041944849

14. Mel Slater. 2009. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 364, 1535 (2009), 3549–3557. DOI: http://dx.doi.org/10.1098/rstb.2009.0138

15. Mel Slater and Maia Garau. 2007. The Use of Questionnaire Data in Presence Studies: Do Not Seriously Likert. *Presence: Teleoperators and Virtual Environments* 16, 4 (2007), 447–456. DOI: http://dx.doi.org/10.1162/pres.16.4.447

16. 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/10.1162/pres.1994.3.2.130

17. Harvey Smith and Matthias Worch. 2010. What Happened Here? Environmental Storytelling. [Online] Last accessed: 26 November 2018. In *Game Developers Conference*. http://www.worch.com/files/gdc/What_Happened_Here_Web_Notes_Small.pdf

18. Agnieszka Sorokowska, Piotr Sorokowski, Peter Hilpert, Katarzyna Cantarero, Tomasz Frackowiak, and Khodabakhsh Ahmadi et al. 2017. Preferred Interpersonal Distances: A Global Comparison. *Journal of Cross-Cultural Psychology* 48, 4 (2017), 577–592. DOI:http://dx.doi.org/10.1177/0022022117698039

19. The Steel Minions. 2018. *The Chantry*. Game [PS4 Digital Download]. (17 September 2018). Sony Interactive Entertainment Europe, London, United Kingdom. Last played September 2018.

20. Stefan Thie and J Wijk. 1998. A general theory on presence: Experimental evaluation of social virtual presence in a decision making task. In *Presence in Shared Virtual Environments Workshop*, Vol. 1.

21. UQO Cyberpsychology Lab. 2004a. Revised Immersive Tendencies Questionnaire 3.01. [Online] Last accessed: 26 November 2018. (2004). http://w3.uqo.ca/cyberpsy/docs/qaires/immersion/ITQ_va.pdf

22. UQO Cyberpsychology Lab. 2004b. Revised Presence Questionnaire 3.0. [Online] Last accessed: 26 November 2018. (2004). http://w3.uqo.ca/cyberpsy/docs/qaires/pres/PQ_va.pdf

23. B. G. Witmer and M. J. Singer. 1998. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence* 7, 3 (June 1998), 225–240. DOI:http://dx.doi.org/10.1162/105474698565686