PRESENCE AND BEYOND: EVALUATING USER EXPERIENCE IN AR/MR/VR

EDITED BY: Richard Skarbez, Missie Smith, Amela Sadagic and Mary C. Whitton
PUBLISHED IN: Frontiers in Virtual Reality

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-83250-139-9
DOI 10.3389/978-2-83250-139-9

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

PRESENCE AND BEYOND: EVALUATING USER EXPERIENCE IN AR/MR/VR

Topic Editors:
Richard Skarbez, La Trobe University, Australia
Missie Smith, Facebook Reality Labs Research (United States), United States
Amela Sadagic, Naval Postgraduate School, United States
Mary C. Whitton, University of North Carolina at Chapel Hill, United States

Citation: Skarbez, R., Smith, M., Sadagic, A., Whitton, M. C., eds. (2022). Presence and Beyond: Evaluating User Experience in AR/MR/VR.
Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83250-139-9

Table of Contents

frontiers | Frontiers in Virtual Reality

TYPE Editorial
PUBLISHED 23 August 2022
DOI 10.3389/frvir.2022.983694

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
George Papagiannakis,
University of Crete, Greece

*CORRESPONDENCE
Richard Skarbez,
skarbez@gmail.com

SPECIALTY SECTION
This article was submitted to Virtual Reality and Human Behavior, a section of the journal Frontiers in Virtual Reality

RECEIVED 01 July 2022
ACCEPTED 19 July 2022
PUBLISHED 23 August 2022

CITATION
Skarbez R, Smith M, Sadagic A and Whitton MC (2022), Editorial: Presence and beyond: Evaluating user experience in AR/MR/VR.
Front. Virtual Real. 3:983694.
doi: 10.3389/frvir.2022.983694

Editorial: Presence and beyond: Evaluating user experience in AR/MR/VR

Richard Skarbez[1]*, Missie Smith[2], Amela Sadagic[3] and Mary C. Whitton[4]

[1]La Trobe University, Melbourne, VIC, Australia, [2]Reality Labs Research, Redmond, WA, United States, [3]Naval Postgraduate School, Monterey, CA, United States, [4]University of North Carolina at Chapel Hill (retired), Chapel Hill, NC, United States

KEYWORDS

augmented reality, virtual reality, mixed reality, presence, user experience

Editorial on the Research Topic
Presence and beyond: Evaluating user experience in AR/MR/VR

1 Introduction

The call for this Research Topic was intentionally broad: We sought papers that identify or propose constructs that can be used to describe AR/MR/VR, and papers that evaluate the utility of those constructs; we sought papers that discussed measures relating to user experience in AR/MR/VR - including, but not limited to, presence. In the end, we were very happy to publish fifteen articles addressing a variety of these questions - but, notably, not all of them. In the remainder of this editorial, we briefly introduce each of the fifteen articles, loosely grouping them into relevant categories. We then discuss each of the three categories in turn, and close with a call to action for our AR/MR/VR research community to more actively engage with human-computer interaction (HCI) and user experience (UX) researchers.

2 Paper summaries

The subsections that follow reflect loose topic categories that will be revisited in Section 3. That said, several articles resisted easy categorization, including these first two.

Ratan and colleagues examine the stereotype threat effect - that is, the fear of behaving in a manner stereotypically associated with one's social group - in the context of VR and AR STEM-gaming applications. Their results suggest that VR and AR experiences may produce different levels of stereotype threat (or its opposite, stereotype reactance).

Neidhardt and Zerlik examine the plausibility of an auditory augmented reality environment that includes position-dynamic binaural synthesis. The subjects wore headphones and could move around independently. The results suggest that inexperienced listeners report a plausible illusion of the spatialized sound; however, the same results did not hold for the experienced listeners.

2.1 Theory

Skarbez, Smith, and Whitton reflect upon the reality-virtuality continuum of Milgram and Kishino. They make several arguments regarding the definition and nature of mixed reality, as well as the continuum itself. For example, they argue that virtual reality - in its present realization - should be considered a subset of MR.

Weinrich and colleagues extensively discuss the nature of the presence construct in Mixed Reality. In the process of doing so, they also propose a modified reality-virtuality continuum and offer a suite of research desiderata and research questions regarding reference frames, transportation, and realism in MR.

Latoschik and Wienrich propose a new model describing experiences across the xR spectrum which takes as its essential conditions congruence (an ontological specification of coherence) and plausibility, from which the place and plausibility illusions can be derived.

Jung and Lindeman present a model for describing the quality of a VR experience using three orthogonal dimensions: coherence, immersion, and illusion; they use *illusion* as an umbrella term for presence and its kin. They go on to argue that user preference is an appropriate metric for evaluating VR experiences.

Hartmann and Hofer propose a psychological parallel processing explanation for users' experiences in xR environments. Their account claims that sensations such as presence are accompanied by the belief that "this is not really happening," which they refer to as media awareness.

Vindenes and Wasson present a post-phenomenological framework for understanding VR experiences, which is to say they propose to study VR as a technology that mediates a human user's relationship with the world.

2.2 Measures

Halbig and Latoschik survey the use of physiological measurements to evaluate virtual reality. They summarize research areas that have used physiological measures and provide tables enumerating the sensors and analysis tools currently available to researchers. We believe this is an excellent and comprehensive resource for researchers.

Hayes, Hughes, and Bailenson report the rigorous initial development of a system of behavioral coding to measure social presence. They validate with a user study and propose directions for future refinement of the system.

2.3 Applications

2.3.1 Social Presence

Miller and Bailenson compare the social presence engendered by virtual humans within the augmented field-of-view and outside it; that is, visible or not visible to the user. The results suggest that users feel less social presence with virtual humans they cannot see.

Sun and Won examine participants' ability to accurately judge one another's emotional state in VR. Participants were represented either as photorealistic or abstract (cube) avatars; the results suggest that participants could correctly judge each other's emotional state regardless of the avatar condition.

(The article by Hayes, Hughes, and Bailenson could have been placed here as well.)

2.3.2 Learning

Bagher and colleagues examine the sense of presence and bodily engagement and their roles in enhancing learners' experience and performance in the context of interactive virtual learning environments. They identify a positive correlation between knowledge gain and the sense of agency supported by embodied affordances.

Ochs and Sonderegger use an experimental mixed-methods approach to evaluate human performance in a memorization task. While participants who learned in VR reported higher levels of presence, participants who learned on a conventional desktop configuration demonstrated better performance on the memorization task.

Carnell and colleagues report on their experience applying the Kirkpatrick Model of training evaluation to medical communication skills training. The results of their study suggest that human behaviors observed in a virtual environment may provide early indicators of how an individual will behave in a comparable real-world scenario.

3 Themes and commonalities

3.1 Theory

A plurality of our published articles present models or frameworks for the description and analysis of AR/MR/VR experiences. Notably, two articles - by Skarbez, Smith, and Whitton and Weinrich and colleagues—propose to modify or extend Milgram and Kishino's reality-virtuality continuum, and two more - by Latoschik and Wienrich and Jung and Lindeman—propose new models that incorporate coherence (congruence in Latoschik and Wienrich) as a key component of their models. These recurring themes communicate the enduring power of these concepts, while simultaneously indicating that they may need to be adapted to suit an evolving technological landscape.

3.2 Measures

Historically, researchers have employed questionnaires and measures of task performance, but the papers published herein highlight the utility of other techniques, such as physiological measurement and behavioral coding. Moving forward, the evaluation of AR/MR/VR systems cannot be limited to single measures, and researchers should triangulate using multiple measures informed by the specific goals of the research and objectives that AR/MR/VR systems are set to support. Rather than evaluating system hardware or software, researchers should aim to evaluate their participants' learning, behavior, and experience.

3.3 Applications

This category includes papers that incorporated user studies primarily focused on social presence and learning applications.

3.3.1 Social presence

AR/MR/VR usage is increasingly social; as such, future research needs to consider not only the individual user's experience of a system, but perhaps the social and cultural effects associated with that system as well. We believe that looking to our colleagues in the social sciences for inspiration, methods, and measures will be a fruitful endeavor for a field that has historically been led by computing scientists and engineers.

3.3.2 Learning

Learning, knowledge, and skill acquisition have always been key areas of AR/MR/VR research, and the articles in this Research Topic reflect that. These results suggest that while AR/MR/VR technologies are exciting new learning tools, they may not be best suited for every learning task; it is important to bear in mind that effective learning can result from any of a number of methods, many of which have been well-studied in related domains. AR/MR/VR learning applications can often benefit from the adoption of best practices from education literature focused on non-immersive learning solutions. The article by Carnell and colleagues is a good example of this: virtual humans are used not as a substitute for, but as an adjunct to, traditional learning structured around the Kirkpatrick model of evaluation for training and learning interventions.

4 Conclusion

In reviewing the articles included in this research topic, we note that authors admirably addressed presence—and social presence—from a variety of perspectives. This aspect of the research topic, then, was clearly a success. That said, many of the specific questions that we raised in the call were not addressed by any of the received manuscripts; there remains ample opportunity for future work in this area.

It may be meaningful that none of these articles adopted the language of "user experience" (UX), nor did they refer to "human-computer interaction" (HCI). We interpret this to signify an unfortunate (in our opinion) siloing of the AR/MR/VR research community. While many issues arise in the study of immersive technologies that are unique to this field, it is just as certain that this work falls within the larger HCI domain - or the UX domain, to use the language preferred by industry. We believe that a failure to situate our field within - and to engage more deeply with - these communities will limit the growth and impact of AR/MR/VR research.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work, and approved it for publication.

Acknowledgments

The authors thank Douglas A. Bowman (Virginia Tech, US) for serving as Editor and shepherd for a paper in this collection for which all of the topic editors declared a conflict of interest.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Virtual Reality

PERSPECTIVE
published: 24 March 2021
doi: 10.3389/frvir.2021.647997

Revisiting Milgram and Kishino's Reality-Virtuality Continuum

Richard Skarbez [1†], Missie Smith [2†] and Mary C. Whitton [3]*

[1] Department of Computer Science and Information Technology, La Trobe University, Melbourne, VIC, Australia, [2] Independent Researcher, Detroit, MI, United States, [3] Department of Computer Science, University of North Carolina, Chapel Hill, NC, United States

Since its introduction in 1994, Milgram and Kishino's reality-virtuality (RV) continuum has been used to frame virtual and augmented reality research and development. While originally, the RV continuum and the three dimensions of the supporting taxonomy (extent of world knowledge, reproduction fidelity, and extent of presence metaphor) were intended to characterize the capabilities of visual display technology, researchers have embraced the RV continuum while largely ignoring the taxonomy. Considering the leaps in technology made over the last 25 years, revisiting the RV continuum and taxonomy is timely. In reexamining Milgram and Kishino's ideas, we realized, first, that the RV continuum is actually discontinuous; perfect virtual reality cannot be reached. Secondly, mixed reality is broader than previously believed, and, in fact, encompasses conventional virtual reality experiences. Finally, our revised taxonomy adds coherence, accounting for the role of users, which is critical to assessing modern mixed reality experiences. The 3D space created by our taxonomy incorporates familiar constructs such as presence and immersion, and also proposes new constructs that may be important as mixed reality technology matures.

Keywords: virtual reality, augmented reality, mixed reality, presence, immersion, coherence, taxonomy

OPEN ACCESS

Edited by:
David Swapp, University College London, United Kingdom

Reviewed by:
Mark Billinghurst, University of South Australia, Australia
Dieter Schmalstieg, Graz University of Technology, Austria

****Correspondence:***
Richard Skarbez
r.skarbez@latrobe.edu.au

† These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Virtual Reality and Human Behaviour, a section of the journal Frontiers in Virtual Reality

Received: *30 December 2020*
Accepted: *05 March 2021*
Published: *24 March 2021*

Citation:
Skarbez R, Smith M and Whitton MC (2021) Revisiting Milgram and Kishino's Reality-Virtuality Continuum. Front. Virtual Real. 2:647997. doi: 10.3389/frvir.2021.647997

1. INTRODUCTION

In 1994, Paul Milgram and Fumio Kishino published "A Taxonomy of Mixed Reality Visual Displays," simultaneously introducing to the literature the notion of the reality-virtuality (RV) continuum and the term "mixed reality" (MR) (Milgram and Kishino, 1994). In the succeeding quarter century, this work has been cited thousands of times, cementing it as one of the seminal works in our field (A related paper by Milgram, Haruo Takemura, Akira Utsumi, and Kishino titled, "Augmented reality: A class of displays on the reality-virtuality continuum," appeared later that year, and also has thousands of citations, Milgram et al., 1994). In that same quarter century, our field has rapidly evolved. Films like *Minority Report*, *Iron Man*, and *Ready Player One* have firmly established augmented reality (AR) and virtual reality (VR) in popular culture. At the same time, AR and VR technologies have rapidly become higher quality, cheaper, and more widely available. As a result, millions of consumers now have access to AR experiences on their mobile phones (e.g., *Pokémon GO*), or VR experiences on the Facebook Oculus or HTC Vive head-mounted displays (e.g., *Beat Saber*). In light of this rapid technological evolution, we believe it is worth revisiting core concepts such as the reality-virtuality continuum.

In this article, we reflect on the RV continuum, the meaning of "mixed reality," and Milgram and Kishino's taxonomy of MR display devices. That reflection leads us to three main points. First, we argue that the RV continuum is, in fact, discontinuous: that the "virtual reality" endpoint is unreachable, and any form of technology-mediated reality is, in fact, mixed reality. Second, we consider the term "mixed reality," and argue for the continuing utility of Milgram and Kishino's definition, with one small but significant change: Instead of requiring that real and virtual objects be combined within a single display, we propose that real and virtual objects and stimuli could be combined within a single percept. Finally, we present a taxonomy—inspired by the taxonomy in Milgram and Kishino's original paper—that can categorize users' mixed reality experiences, and discuss some of the implications of this taxonomy.

We choose to focus our discussion specifically on the concepts and constructs introduced by Milgram and Kishino in their original papers. We make this choice in the interest of clarity and readability. However, we acknowledge that we are not the first or the only authors to expand upon their work in the last 20 years. Koleva, Benford, and Greenhalgh explored the idea of *boundaries* between physical and virtual spaces in mixed reality environments, and delineated some of their properties (Koleva et al., 1999). Lindeman and Nova proposed a classification framework for multisensory AR experiences based on where the real and virtual stimuli are mixed (Lindeman and Noma, 2007). Normand, Servières, and Moreau reviewed existing taxonomies of AR applications and proposed their own (Normand et al., 2012). Barba, MacIntyre, and Mynatt argued for using a definition of MR inspired by Mackay (Mackay, 1998) and a definition of AR from Azuma (Azuma, 1997), and used the RV continuum to describe the relationship between the two (Barba et al., 2012). Mann and colleagues discussed a variety of "realities"–virtual, augmented, mixed, and mediated–and proposed *multimediated* reality (Mann et al., 2018). Speiginer and MacIntyre introduced the concept of *reality layers* and proposed the *Environment-Augmentation framework* for reasoning about mixed reality applications (Speiginer and MacIntyre, 2018). Speicher, Hall, and Nebeling investigated the definition of mixed reality through a literature review and a series of interviews with domain experts. They also proposed a conceptual framework for MR (Speicher et al., 2019). Of note is that while these papers enrich the discussion regarding mixed reality, none challenges the central notion of the RV continuum, nor do they generally propose alternative definitions of mixed reality.

2. REVISITING THE REALITY-VIRTUALITY CONTINUUM

The RV continuum, as initially proposed by Milgram and Kishino, is shown in **Figure 1**. They anchor one end with a purely real environment, "consisting solely of real objects," and the other, with a purely virtual environment, "consisting solely of virtual objects" (Milgram and Kishino, 1994). They consider any environment which consists of a blending of real and virtual objects to be mixed reality (MR). Mixed reality environments where the real world is augmented with virtual content are called augmented reality (AR), while those where most of the content is virtual but there is some awareness or inclusion of real world objects are called augmented virtuality (AV). Of note is that this original version of the continuum was explicitly concerned only with visual displays.

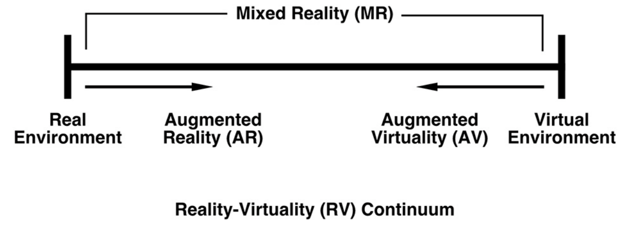

FIGURE 1 | Milgram and Kishino's reality-virtuality continuum (adapted from Milgram et al., 1994).

While the original version of the continuum has undoubtedly served the field well, we have identified limitations. One is that, as mentioned above, Milgram and Kishino were explicitly concerned with visual displays, and primarily with display hardware. A second is that nowhere in this continuum is seen the notion of an observer or a user with senses other than visual and prior life experiences. Finally, content was described only in relation to realism (e.g., wireframes vs. 3D renderings), with no concern for the coherence of the overall experience. We will soon argue that the notion of an environment without an experiencing being—the aforementioned observer—is incomplete. That is, the mediating technology, content conveyed, and resulting user experience must be considered together to adequately describe MR experiences.

The first limitation is fairly straightforward, and in fact was commented upon by Milgram and Kishino in their original paper: "It is important to point out that, although we focus in this paper exclusively on mixed reality visual displays, many of the concepts proposed here pertain as well to analogous issues associated with other display modalities" (Milgram and Kishino, 1994). In our revisiting of the RV continuum we have taken into consideration the advances in synthesizing and displaying data for the multiple senses.

Today's processor speeds make it possible to deliver high quality *audio signals*, for instance, by modeling room acoustics (Savioja and Svensson, 2015) and sound propagation in multi-room spaces (Liu and Manocha, 2020).

Haptic displays mimic solid surfaces and other tactile stimuli. Haptics can be *active*, with solid surfaces approximated with forces supplied by a device (Salisbury and Srinivasan, 1997), or *passive* where the user feels real objects that correspond to virtual objects (Insko, 2001; Azmandian et al., 2016).

Heilig's 1962 Sensorama (Heilig, 1962) presaged integration of *scent* into virtual reality systems (Yanagida, 2012). Olfactory interfaces have matured to the point that recent work by Flavián, Ibáñez-Sánchez, and Orús explored how to make olfactory input *more effective*, rather than simply focusing on making it work (Flavián et al., 2021).

The complete *taste* experience combines sound, smell, haptics, and a chemical substance that mimics natural taste and simulates the taste buds. The Food Simulator project (Iwata et al., 2004) tackled the haptic component of taste, and recent work reports on a taste display that synthesizes and delivers tastes that match those sampled with a taste sensor (Miyashita, 2020).

All that is to say that researchers have now demonstrated at least preliminary abilities to deliver computer-generated stimuli to all the exteroceptive senses—that is, those senses responding to stimuli that come from outside the body. As progress continues, we may approach the capabilities of Ivan Sutherland's Ultimate Display—"a room within which the computer can control the existence of matter" (Sutherland, 1965) (In popular culture, one can see the Ultimate Display in the Holodecks of the *Star Trek* franchise).

However, we argue that even if we were to have the Ultimate Display, it would still fall within the realm of mixed reality. That is because, even with total control of a user's exteroceptive senses, we still would not have control over their *interoceptive senses*—the senses that monitor the body's internal state, such as the vestibular and proprioceptive senses. And even in the Ultimate Display, there would be circumstances in which these interoceptive senses would be in conflict with the information being supplied to the exteroceptive senses. For example, consider how you, as a user, might feel if the Ultimate Display were used to generate a virtual environment depicting a spacewalk. The visual display could be completely indistinguishable from the real thing, but you would still know which direction *down* was, and your feet would still be on the floor. We argue that these sensory conflicts are *inherent* to conventional virtual reality systems (which we refer to as *External Virtual Environments* in **Figure 2**). Observers experience external virtual environments through stimulation of the five basic exteroceptive senses (i.e., sight, hearing, touch, smell, and taste) while interoceptive senses remain unaltered. An important characteristic of external virtual environments is that they are *unable* to manipulate interoceptive senses.

There is, however, a popular conception of a "virtual environment" in which these sensory conflicts could be avoided: the Matrix, from the popular film series of the same name. In the *Matrix* films, sensory agreement is accomplished by direct brain stimulation: a person's sensory organs are in some way disconnected from their brain such that both interoceptive (e.g., proprioception) and exteroceptive (e.g., sight) senses are stimulated by technology. We argue that this is the only type of virtual environment that could exist outside of the mixed reality spectrum. Every other system, even the Ultimate Display, presents mixed—and potentially conflicting—exteroceptive and/or interoceptive stimuli to the user. Following this logic, we present our revised RV continuum seen in **Figure 2**, which, on the right end, includes a discontinuity between external virtual environments and the right-end anchor, "Matrix-like" VR.

We feel that the virtual environment endpoint in the original continuum was ill-defined, being any environment "consisting solely of virtual objects," although it was implied that such an environment "is one in which the participant-observer is totally immersed in, and able to interact with, a completely synthetic world" (Milgram and Kishino, 1994). Most subsequent authors seem to have assumed that the virtual environment endpoint comprises what we have called external VEs, but, as we have argued, these are never "totally immers[ive]" or "completely synthetic" because they cannot control or manipulate the interoceptive senses. Furthermore, the display devices in such external VEs *are themselves real objects*, situated in the real environment. As a result, users experience such external VEs as mixed reality, with virtual objects situated within a real environment. The discontinuity in our revised continuum makes it explicit that there are real and substantial differences between external virtual environments and "Matrix-like" virtual environments.

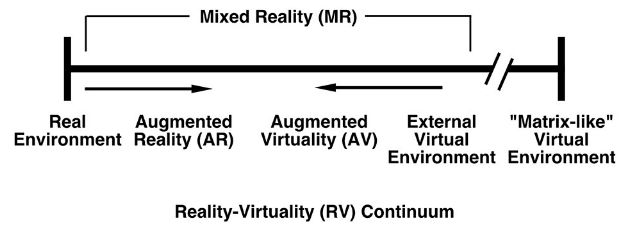

FIGURE 2 | Our revised reality-virtuality continuum. Note that the External Virtual Environment (traditionally called "Virtual Reality") is still part of MR.

3. THE MEANING OF MIXED REALITY

> "Within this [reality-virtuality] framework it is straightforward to define a generic Mixed Reality (MR) environment as one in which real world and virtual world objects are presented together within a single display" (Milgram et al., 1994).

The preceding quote clearly defines MR, at least as Milgram and his colleagues envisioned it. MR is any display (interpreted broadly) that presents a combination of real and virtual objects that are perceived at the same time. This can be achieved in a variety of ways. Virtual objects can be visually overlaid on the real world, using optical- or video-see-through display techniques. Alternatively, real world content can be integrated into a virtual world by embedding a live video stream or, appealing to a different sense, by incorporating tracked haptic objects into a virtual experience.

Since Milgram and Kishino's initial publication, researchers have arrived at vastly different and sometimes conflicting definitions of MR. For example, MR has been defined as a combination of AR and VR, as a synonym for AR, as a "stronger" version of AR, or as Milgram and Kishino defined it (Speicher et al., 2019). In popular culture, the distinction between augmented and mixed reality has also been blurred, with some companies such as Intel[1] describing mixed reality as spatially-located and interactive with the real world, while augmented reality specifically does not include interaction. Microsoft[2] defines augmented reality as the overlaying of graphics onto video—such as AR presented on mobile phones or tablets—while mixed reality requires a combination of the physical and the

[1]https://www.intel.com/content/www/us/en/tech-tips-and-tricks/virtual-reality-vs-augmented-reality.html
[2]https://docs.microsoft.com/en-us/windows/mixed-reality/discover/mixed-reality

virtual. An example is the Microsoft HoloLens game RoboRaid[3], in which enemies seem to exist *on* the walls and can be occluded by real objects in the real room in which the game is being played; if you move to a different room, the enemies' locations adapt to the new physical configuration. A commonly-employed shorthand is that MR systems possess knowledge about the physical world, while AR systems do not (The notion of world knowledge is discussed at length in the following section).

We propose to unify these various definitions by making a small but fundamental change to Milgram et al.'s original definition of mixed reality. This change addresses the second limitation noted in section 2, i.e., the missing user/observer. To account for the importance of how the real or virtual content is *observed*, we propose this definition: a mixed reality (MR) environment is one in which real world and virtual world objects and stimuli are presented together within a single *percept*. That is, when a user simultaneously perceives both real and virtual content, *including across different senses*, that user is experiencing mixed reality. As such, our definition agrees with Milgram et al.'s original assertion that augmented reality is a subset of mixed reality. However, we argue that external virtual reality, what some consider to be the end point of the original RV continuum, is *also* a subset of mixed reality, because an individual may perceive virtual content with some senses and real content with others (including interoception). For example, simulating eating a meal by applying the most sophisticated visual, audio, haptic, olfactory, and taste cues may be convincing to a user, but at some point they would likely realize that they are still not satiated, and in fact, may be more hungry than when they began. This conflict between exteroception and interoception shows how conflicting signals in a single percept can make an experience incongruent. It is for this reason that there is a discontinuity on our revised continuum, because true virtual reality exists only when all senses—exteroceptive and interoceptive—are fully overridden by computer-generated content.

We acknowledge the potential criticism that our broaddefinition that includes external virtual environments makes "mixed reality" too inclusive and potentially confusing. (In the words of Speicher, Hall, and Nebeling's interviewees, "if [a console video game] is MR, then everything is" Speicher et al., 2019). Milgram and Kishino's original definition required the *(visually) displayed content* to be a mixture of real and virtual, while our proposed redefinition merely requires the user's *overall sensory experience, the percept,* be a mixture of real and virtual. Our response to the criticism that our definition of MR is too broad is two-fold. First, as illustrated earlier in this section, the many definitions of MR were already a source of confusion. Second, as we discuss in the following section, "mixed reality" is not intended to fully describe a system or an experience. This was clear in Milgram and Kishino's conception as well: They supplemented their RV continuum with their less well-known taxonomy for characterizing mixed reality technology. In the next section, we revisit Milgram and Kishino's taxonomy and propose an updated version.

[3]https://www.microsoft.com/en-us/p/roboraid/9nblggh5fv3j

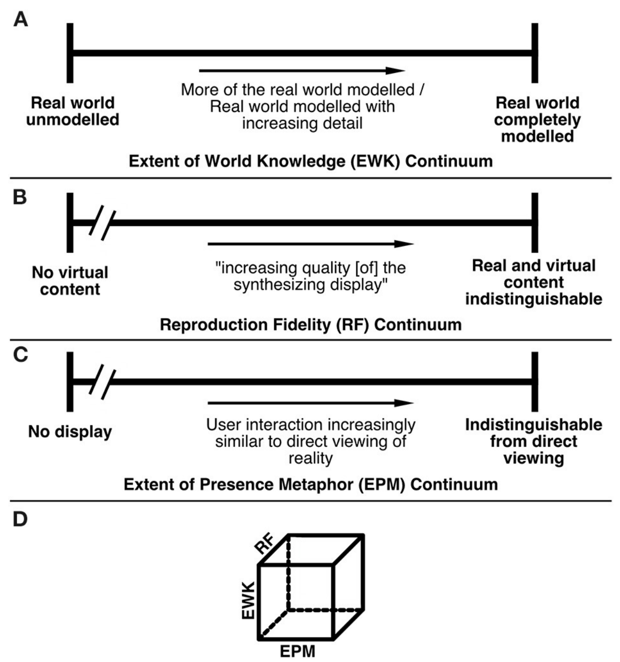

FIGURE 3 | The EWK **(A)**, RF **(B)**, and EPM **(C)** continuua, as well as the proposed framework combining the three **(D)** (adapted from Milgram and Kishino, 1994; Milgram et al., 1994).

4. A NEW TAXONOMY OF MR EXPERIENCES

4.1. Milgram and Kishino's Original Taxonomy of MR *Systems*

Milgram and Kishino's original paper included a three-dimensional taxonomy to characterize various mixed reality technologies. First, **extent of world knowledge (EWK)** described the level of modeling of the real world, and specifically the *where* (locations of objects) or *what* (identification of objects), included in the MR environment. Second, the **reproduction fidelity (RF)** of a technology described a display technology's capability of exactly reproducing the real world. Finally, the **extent of presence metaphor (EPM)** accounted for the level of world-conformal graphics and viewpoint of the person experiencing the MR environment (Essentially, the naturalness of the user's interaction with the display). These dimensions, and how Milgram and Kishino viewed their relationship, can be seen in **Figure 3**.

4.2. Our Proposed Taxonomy for MR *Experiences*

In the mid-1990s, both head-worn displays and computer hardware were generally bulky and, except for a few systems such as the Sony Glasstron and Virtual i-O i-glasses!, expensive. A typical research laboratory system minimally required a head-worn display, a high-performance workstation, and a tracking system. Total system cost could easily exceed 100,000

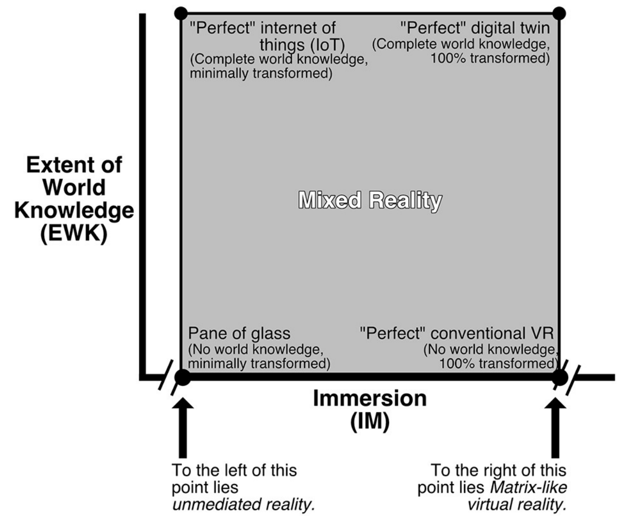

FIGURE 4 | The 2D space defined by Immersion (IM) and Extent of World Knowledge (EWK), with examples.

USD. Except for demonstration programs and a few games, most applications were custom developed, often on custom or customized hardware. Examples include Disney's virtual reality application *Aladdin* (Pausch et al., 1996), State et al.'s augmented reality system for ultrasound-guided needle biopsies (State et al., 1996), and Feiner et al.'s augmented reality application *Touring Machine* (Feiner et al., 1997).

In 2020, most mixed reality experiences can be implemented using off-the-shelf hardware solutions for components such as visual displays, processors, trackers, and user input devices. Even off-the-shelf display devices (including mobile phones or tablets used as AR displays, head-worn AR displays, and head-worn VR displays) often include tracking in addition to integrated processing. Therefore, except for custom systems, the main differentiator among MR experiences at near points on the continuum is no longer the mediating technology, but instead is the user's overall experience (This notion is echoed by Speicher, Hall, and Nebeling's interviewees: "[I]n the future, we might distinguish based on applications rather than technology," Speicher et al., 2019). In response to the major technology changes since the mid 1990s, we propose to modify and expand Milgram and Kishino's taxonomy in order to be able use it to categorize not only mixed reality technologies, but also, and importantly, mixed reality *experiences*.

Two of our proposed dimensions derive from Milgram and Kishino's original three. We adopt the **Extent of World Knowledge (EWK)** dimension directly, as we feel this captures a key component of augmented reality and augmented virtuality experiences—the extent to which the system is aware of its real world surroundings and can respond to changes in those surroundings. While, Milgram described EWK as a combination of *what* and *where* known objects are, modern sensing technologies, such as imagined in a pervasive *internet of things*, could provide access to much richer streams of information about the real world environment. "Perfect" EWK would take advantage of these additional sensing capabilities wherever available and would extend to *how* things work, and *when* things might happen.

We propose to combine the Reproduction Fidelity (RF) and Extent of Presence Metaphor (EPM) dimensions into a single dimension, **Immersion (IM)**. We adopt the name following the definition of immersion favored by Slater et al. (Slater, 2004; Skarbez et al., 2017). That is, a system's immersion is the set of valid actions supported by that system. We choose to combine RF and EPM based on similarities between these dimensions, hinted at by **Figure 3** and remarked upon by Milgram and Kishino in the original paper: "…so too is the EPM axis in some sense not entirely orthogonal to the RF axis, since each dimension independently tends toward an extremum which ideally is indistinguishable from viewing reality directly" (Milgram and Kishino, 1994). Both the RF and EPM dimensions have a discontinuity at the minimum, as, when there is no display, the real world is perceived in an unmediated fashion, which, by definition, is indistinguishable from reality. We argue that this similarity in the two dimensions is not an accident; a system's immersion has the same behavior.

Furthermore, Slater has argued that immersion comprises two types of valid actions: *sensorimotor* valid actions and *effectual* valid actions. These are valid actions that result in changes to a user's perception of the environment and changes to the environment itself, respectively. We argue that both RF and EPM are actually part of sensorimotor valid actions. The original Milgram and Kishino paper does not account for the possibility of a system being interactive, beyond the choice of viewpoint. This limitation is also addressed by the inclusion of effectual valid actions in the combined dimension.

Already, with these two dimensions, IM and EWK, we can begin to productively characterize MR systems (**Figure 4**). External virtual reality systems generally have high immersion, with little or no world knowledge. Augmented reality systems, on the other hand, have low or medium immersion, but a higher level of world knowledge. A 2D mapping of these dimensions would show four extremes with IM along the x-axis and EWK on the y-axis. At the bottom left, a *pane of glass* represents no world knowledge or immersion at all. Perfect world knowledge with no immersion could be thought of as an *internet of things (IoT)* system, wherein the system knows the state of the real world, but does not itself display that state to a user. A fully immersive system with no world knowledge is the ideal of *conventional VR*, in which the virtual world is rendered exquisitely, but the system does not consider the real environment. The top right corner of the graph requires high IM and high EWK together, which results in a perfect *digital twin*, offering real-time tracking and rendering of the real world.

Populating this entire 2D IM-EWK space with examples is beyond the scope of this paper, as there are far too many to include. However, the upper-right portion of the space, near the extrema that we have labeled as a "perfect" digital twin, has not been as widely explored. One thread of research in this space is MR telepresence, in which the goal is to capture remote spaces and users and reproduce them elsewhere in real time. Recent

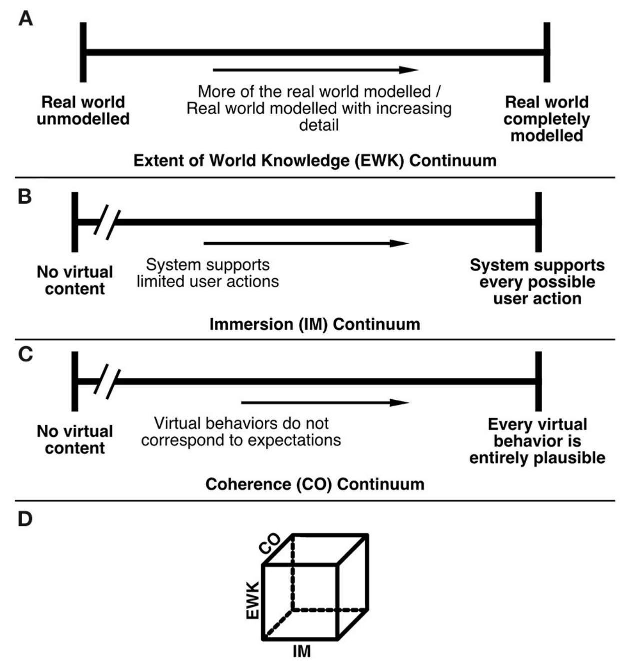

FIGURE 5 | Our three dimensional taxonomy consisting of EWK **(A)**, IM **(B)**, and CO **(C)** dimensions, as well as the relationship among the three **(D)**.

papers in this area include Kunert et al. (2018), Stotko et al. (2019b), and Stotko et al. (2019a). Another research thread is what we call world-aware environment generation, in which the goal is to create a virtual environment that possesses some of the same characteristics of the real environment—for example, areas that are not navigable in the real environment are also not navigable in the virtual environment. Recent research in this area includes Sra et al. (2016) and Cheng et al. (2019).

To describe the essence of a user's experience, and to address the third limitation identified in section 2, we must go beyond the elements of realism contained in EWK and IM to consider how sensory inputs create a unified experience–the coherence of the experience. We argue that this third dimension already exists in the literature. Various authors refer to it alternatively as *fidelity* (Alexander et al., 2005) or authenticity (Gilbert, 2017), but we prefer the term **Coherence (CO)**, following (Skarbez et al., 2017). Our resulting taxonomy is illustrated in **Figure 5**. While, EWK and IM can be thought of as describing what the system is *intended to do*, CO describes *how consistently that intention is conveyed to the user*. In the context of VR, coherence is primarily *internal*; that is, do virtual objects interact with one another and the user in predictable ways? In the context of AR, coherence is primarily *external*; that is, do virtual objects interact with real objects and the user in predictable ways? For example, in AR, objects meant to be fixed in the world would be externally incoherent if they float in space rather than sitting on a surface. In applications that are "truly" MR, both internal and external coherence are necessary for a satisfying and effective user experience.

4.3. Implications of Our Revised Taxonomy

In the remainder of this section, we will take it as given that our proposed taxonomy is appropriate—that it consists of three logically distinct, if not wholly orthogonal dimensions, and that these dimensions span a meaningful portion of the space of possible MR experiences. Each of the following subsections addresses a consequence which derives from this taxonomy.

4.3.1. Appropriate Constructs for Describing MR Experiences

We believe that Extent of World Knowledge, Immersion, and Coherence are objective characteristics of MR systems and that metrics could be identified for each of them. A variety of subjective constructs have been proposed to evaluate a user's experience of VR and AR systems, perhaps the most well-known of which is *presence*. Elsewhere in the literature, it has been argued that presence results from the combination of *Place Illusion (PI)*—otherwise known as *spatial presence*—and *Plausibility Illusion (Psi)*, with Place Illusion arising from the immersion of a system (Slater, 2009), and Plausibility Illusion from the coherence of a system (Skarbez, 2016). Even if one accepts that PI and Psi are appropriate constructs and that one has valid means to measure them—a difficult and contentious topic itself—neither of them contains any notion of extent of world knowledge. To our knowledge, there has been little or no research regarding EWK since Milgram et al.'s original papers on the topic. Certainly no constructs have been proposed that combine EWK with IM and/or CO. This is an area ripe for future research. We do not claim to have all the answers in this area, but in the interest of stimulating discussion, we propose the following model (**Figure 6**; constructs that have not yet been named or discussed in the literature are in *ITALICS*):

Some brief commentary on these new constructs:

- By *world awareness*, we mean a user's feeling that the system is aware of the physical world around them.
- By *replicated world illusion*, we mean a user's feeling that they are in a virtual copy of the real world (which may be analogous to the concept of *telepresence* as described in Steuer, 1992).
- By *system intelligence illusion*, we mean a user's feeling that the system itself is aware of its surroundings and uses that awareness intelligently; that is, in ways that do not violate coherence.
- By *mixed reality illusion*, we mean a user's feeling that they are in a place that blends real and virtual stimuli seamlessly and responds intelligently to user behavior.

These constructs are also useful in describing the discontinuities on the continua. We posit that having no immersion at all, or no computer-generated stimuli, means that MR is not possible. Thus, regardless of the quantity of EWK or coherence present, if immersion is non-existent, then so is the MR experience.

4.3.2. The Difficulty of Construct Measurement

Note that the objective nature of Extent of World Knowledge, Immersion, and Coherence doesn't mean that they are easy to measure. Far from it! It just means that they are characteristics of

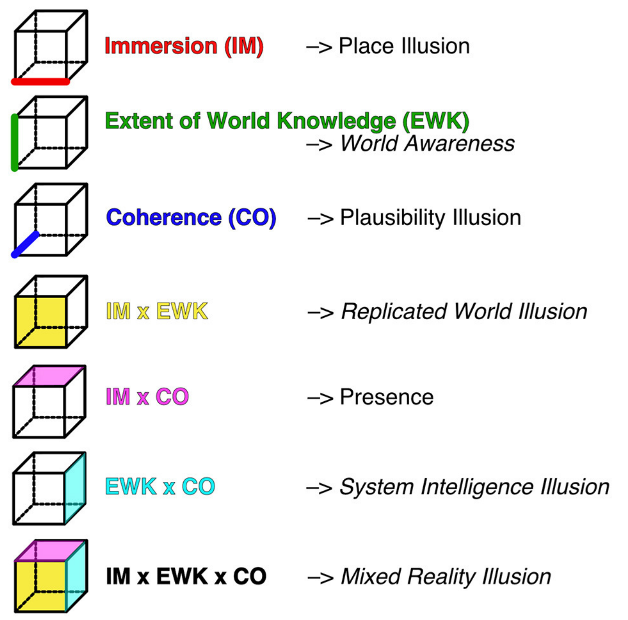

FIGURE 6 | Each cube illustrates a combination of objective system dimensions and the subjective feeling that arises in a user who experiences it.

the *system*, not of the system's *user*. Speicher et al. also developed a conceptual framework for describing objective characteristics of MR, which include: number of environments, number of users, level of immersion, level of virtuality, degree of interaction, and input/output. While these dimensions can help describe a system, they do not provide guidance for evaluating users' experiences. However, by applying our 3D framework to the description of a given MR experience, we believe it may be possible to generalize recommendations for assessing the experience. Assigning specific values for a given system on any of the EWK, IM, or CO axes remains a substantial open research problem. For example, when considering the immersion continuum, should we place technology based on its immersion across all senses, or should each sense be measured in isolation (adding significant complexity as it would change our one-dimensional immersion continuum into six or seven dimensions)? Further, how do we measure how far along in the continuum a technology is? Categorizing MR experiences is difficult, but placing it in our 3D space can guide researchers and practitioners as to what measures (previously or currently used, or perhaps still to be developed) may be most appropriate for evaluating users' MR experiences.

4.3.3. Evaluating MR Experiences

In section 4.3.1, we proposed a set of constructs—some already well-represented in the literature, others described here for the first time—that could be used to describe a user's experience of an MR application. Under some circumstances, these constructs may suffice for evaluation of such applications. In most cases, however, they will not. This is because, when evaluating an application, it is important to consider the intent of its creators. For a virtual reality game, it is more important whether it is *entertaining* than whether it is immersive. For a virtual reality stress induction protocol, it is more important that it is *stressful* and *controllable* than that it gives rise to presence. For an automotive AR head-up display, it is more important that it be *useful* and *safe* than for it to be highly world-aware. All of this is to say that evaluation is a different process than characterization; it requires different, and in many cases *application-specific* measures. However, accurate characterization helps us identify what measures may be most appropriate and valid for a given scenario.

5. CONCLUSION

Our intention with this article has not been to disparage Milgram and Kishino's work; on the contrary, we think it has admirably stood the test of time, and deserves to be recognized as one of the seminal papers in the field. That said, with the benefit of hindsight, there are some areas that we feel needed updating for the concepts to remain relevant in 2020 and beyond. To that end, we proposed a revised version of the reality-virtuality continuum based on the idea that virtual content is always ultimately situated in the real world, which has the consequence that conventional virtual reality should fall within the category of mixed reality. We argued for the continued relevance of a "big tent" definition of MR, and in fact, argued to make the tent bigger still by including all technology-mediated experiences under the term mixed reality. We presented a new taxonomy for describing MR experiences with the dimensions *extent of world knowledge, immersion, and coherence*. The new taxonomy was inspired by Milgram and Kishino's taxonomy which we feel has been underappreciated in comparison to the other contributions made by their original paper. Much as we were inspired by and are indebted to Milgram and Kishino's original work, we hope that this paper encourages further discussions and research in this area.

DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article.

AUTHOR CONTRIBUTIONS

RS, MS, and MW contributed to the ideation and development of this article. MS developed the initial draft, which was substantially revised and expanded by RS. RS, MS, and MW contributed to further revision and polishing of the final article. All authors contributed to the article and approved the submitted version.

REFERENCES

Alexander, A. L., Brunyé, T., Sidman, J., and Weil, S. A. (2005). "From gaming to training: A review of studies on fidelity, immersion, presence, and buy-in and their effects on transfer in PC-based simulations and games," in *The Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)* (Orlando, FL: NTSA).

Azmandian, M., Hancock, M., Benko, H., Ofek, E., and Wilson, A. D. (2016). "Haptic retargeting: dynamic repurposing of passive haptics for enhanced virtual reality experiences," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16* (New York, NY: ACM), 1968–1979. doi: 10.1145/2858036.2858226

Azuma, R. T. (1997). A survey of augmented reality. *Presence Teleoperat Virtual Environ*. 6, 355–385. doi: 10.1162/pres.1997.6.4.355

Barba, E., MacIntyre, B., and Mynatt, E. D. (2012). Here we are! Where are we? Locating mixed reality in the age of the smartphone. *Proc. IEEE* 100, 929–936. doi: 10.1109/JPROC.2011.2182070

Cheng, L., Ofek, E., Holz, C., and Wilson, A. D. (2019). "Vroamer: generating on-the-fly VR experiences while walking inside large, unknown real-world building environments," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (Osaka), 359–366. doi: 10.1109/VR.2019.8798074

Feiner, S., MacIntyre, B., Höllerer, T., and Webster, A. (1997). A touring machine: prototyping 3D mobile augmented reality systems for exploring the urban environment. *Pers. Technol.* 1, 208–217. doi: 10.1007/BF01682023

Flavián, C., Ibáñez Sánchez, S., and Orús, C. (2021). The influence of scent on virtual reality experiences: the role of aroma-content congruence. *J. Bus. Res*. 123, 289–301. doi: 10.1016/j.jbusres.2020.09.036

Gilbert, S. B. (2017). Perceived realism of virtual environments depends on authenticity. *Presence Teleoperat. Virtual Environ*. 25, 322–324. doi: 10.1162/PRES_a_00276

Heilig, M. L. (1962). *Sensorama Simulator*. U.S. Patent No 3,050,870. Washington, DC: U.S. Patent and Trademark Office.

Insko, B. E. (2001). *Passive haptics significantly enhances virtual environments* (Ph.D. thesis). University of North Carolina at Chapel Hill, Chapel Hill, NC, United States.

Iwata, H., Yano, H., Uemura, T., and Moriya, T. (2004). "Food simulator: a haptic interface for biting," in *IEEE Virtual Reality 2004* (Chicago, IL: IEEE), 51–57. doi: 10.1109/VR.2004.1310055

Koleva, B., Benford, S., and Greenhalgh, C. (1999). *The Properties of Mixed Reality Boundaries*. Dordrecht: Springer. doi: 10.1007/978-94-011-4441-4_7

Kunert, C., Schwandt, T., and Broll, W. (2018). "Efficient point cloud rasterization for real time volumetric integration in mixed reality applications," in *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (Munich), 1–9. doi: 10.1109/ISMAR.2018.00023

Lindeman, R. W., and Noma, H. (2007). "A classification scheme for multi-sensory augmented reality," in *Proceedings of the 2007 ACM Symposium on Virtual Reality Software and Technology, VRST '07* (New York, NY: Association for Computing Machinery), 175–178. doi: 10.1145/1315184.1315216

Liu, S., and Manocha, D. (2020). Sound synthesis, propagation, and rendering: a survey. *arXiv [preprint]. arXiv:2011.05538*. Available online at: https://arxiv.org/abs/2011.05538

Mackay, W. E. (1998). "Augmented reality: linking real and virtual worlds: a new paradigm for interacting with computers," in *Proceedings of the Working Conference on Advanced Visual Interfaces, AVI '98* (New York, NY: Association for Computing Machinery), 13–21. doi: 10.1145/948496.948498

Mann, S., Furness, T., Yuan, Y., Iorio, J., and Wang, Z. (2018). All reality: Virtual, augmented, mixed (x), mediated (x, y), and multimediated reality. *arXiv [preprint]. arXiv:1804.08386*. Available online at: https://arxiv.org/abs/1804.08386

Milgram, P., and Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE Trans. Inform. Syst*. 77, 1321–1329.

Milgram, P., Takemura, H., Utsumi, A., and Kishino, F. (1994). Augmented reality: a class of displays on the reality-virtuality continuum. *Proc. SPIE* 2351, 282–292. doi: 10.1117/12.197321

Miyashita, H. (2020). "Norimaki synthesizer: taste display using ion electrophoresis in five gels," in *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI), 1–6. doi: 10.1145/3334480.3382984

Normand, J.-M., Serviéres, M., and Moreau, G. (2012). "A new typology of augmented reality applications," in *Proceedings of the 3rd Augmented Human International Conference, AH '12* (New York, NY: Association for Computing Machinery). doi: 10.1145/2160125.2160143

Pausch, R., Snoddy, J., Taylor, R., Watson, S., and Haseltine, E. (1996). "Disney's aladdin: first steps toward storytelling in virtual reality," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '96* (New York, NY: ACM), 193–203. doi: 10.1145/237170.237257

Salisbury, J. K., and Srinivasan, M. A. (1997). Phantom-based haptic interaction with virtual objects. *IEEE Comput. Graph. Appl.* 17, 6–10. doi: 10.1109/MCG.1997.1626171

Savioja, L., and Svensson, U. P. (2015). Overview of geometrical room acoustic modeling techniques. *J. Acoust. Soc. Am*. 138, 708–730. doi: 10.1121/1.4926438

Skarbez, R. (2016). *Plausibility illusion in virtual environments* (Ph.D. thesis). The University of North Carolina at Chapel Hill, Chapel Hill, NC, United States.

Skarbez, R., Brooks, F. P. Jr., and Whitton, M. C. (2017). A survey of presence and related concepts. *ACM Comput. Surv*. 50, 96:1–96:39. doi: 10.1145/3134301

Slater, M. (2004). "A note on presence terminology," in *Presence Connect*, Vol. 3, 1–5. Available online at: http://www0.cs.ucl.ac.uk/research/vr/Projects/Presencia/ConsortiumPublications/ucl_cs_papers/presence-terminology.htm

Slater, M. (2009). Place illusion and plausibility can lead to realistic behavior in immersive virtual environments. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci*. 364, 3549–3557. doi: 10.1098/rstb.2009.0138

Speicher, M., Hall, B. D., and Nebeling, M. (2019). "What is mixed reality?" in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI '19* (New York, NY: Association for Computing Machinery), 1–15. doi: 10.1145/3290605.3300767

Speiginer, G., and MacIntyre, B. (2018). "Rethinking reality: a layered model of reality for immersive systems," in *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)* (Munich), 328–332. doi: 10.1109/ISMAR-Adjunct.2018.00097

Sra, M., Garrido-Jurado, S., Schmandt, C., and Maes, P. (2016). "Procedurally generated virtual reality from 3D reconstructed physical space," in *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology, VRST '16* (New York, NY: Association for Computing Machinery), 191–200. doi: 10.1145/2993369.2993372

State, A., Livingston, M. A., Garrett, W. F., Hirota, G., Whitton, M. C., Pisano, E. D., et al. (1996). "Technologies for augmented reality systems: realizing ultrasound-guided needle biopsies," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques* (New Orleans, LA), 439–446. doi: 10.1145/237170.237283

Steuer, J. (1992). Defining virtual reality: dimensions determining telepresence. *J. Commun*. 42, 73–93. doi: 10.1111/j.1460-2466.1992.tb00812.x

Stotko, P., Krumpen, S., Klein, R., and Weinmann, M. (2019a). "Towards scalable sharing of immersive live telepresence experiences beyond room-scale based on efficient real-time 3D reconstruction and streaming," in *CVPR Workshop on Computer Vision for Augmented and Virtual Reality* (Long Beach, CA), 3. Available online at: https://xr.cornell.edu/s/7_stotko2019cv4arvr.pdf

Stotko, P., Krumpen, S., Weinmann, M., and Klein, R. (2019b). "Efficient 3D reconstruction and streaming for group-scale multi-client live telepresence," in *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (Beijing), 19–25. doi: 10.1109/ISMAR.2019.00018

Sutherland, I. (1965). "The ultimate display," in *Proceedings of the International Federation of Information Processing (IFIP)* (New York, NY), 506–508.

Yanagida, Y. (2012). "A survey of olfactory displays: making and delivering scents," in *Sensors, 2012* (Taipei: IEEE), 1–4. doi: 10.1109/ICSENS.2012.6411380

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Virtual Reality

HYPOTHESIS AND THEORY
published: 29 April 2021
doi: 10.3389/frvir.2021.656423

A Postphenomenological Framework for Studying User Experience of Immersive Virtual Reality

Joakim Vindenes[1,2]* and *Barbara Wasson*[1,2]

[1] Centre for the Science of Learning & Technology (SLATE), University of Bergen, Bergen, Norway, [2] Department of Information Science and Media Studies, University of Bergen, Bergen, Norway

Virtual Reality (VR) is a remarkably flexible technology for interventions as it allows the construction of virtual worlds with ontologies radically different from the real world. By embodying users in avatars situated in these virtual environments, researchers can effectively intervene and instill positive change in the form of therapy or education, as well as affect a variety of cognitive changes. Due to the capabilities of VR to mediate both the environments in which we are immersed, as well as our embodied, situated relation toward those environments, VR has become a powerful technology for "changing the self." As the virtually mediated experience is what renders these interventions effective, frameworks are needed for describing and analyzing the mediations brought by various virtual world designs. As a step toward a broader understanding of how VR mediates experience, we propose a post-phenomenological framework for describing VR mediation. Postphenomenology is a philosophy of technology concerned with empirical data that understand technologies as mediators of human-world relationships. By addressing how mediations occur *within* VR as a user-environment relation and *outside* VR as a human-world relation, the framework addresses the various constituents of the virtually mediated experience. We demonstrate the framework's capability for describing VR mediations by presenting the results of an analysis of a selected variety of studies that use various user-environment relations to mediate various human-world relations.

Keywords: user experience, virtual reality, postphenomenology, mediation theory, framework

Edited by:
Richard Skarbez,
La Trobe University, Australia

Reviewed by:
Bas De Boer,
University of Twente, Netherlands
Phil Lopes,
École Polytechnique Fédérale de Lausanne, Switzerland

Correspondence:
Joakim Vindenes
joakim.vindenes@uib.no

Specialty section:
This article was submitted to Virtual Reality and Human Behaviour, a section of the journal Frontiers in Virtual Reality

Received: *20 January 2021*
Accepted: *24 March 2021*
Published: *29 April 2021*

Citation:
Vindenes J and Wasson B (2021) A Postphenomenological Framework for Studying User Experience of Immersive Virtual Reality. Front. Virtual Real. 2:656423. doi: 10.3389/frvir.2021.656423

INTRODUCTION

VR technologies are valuable and versatile tools because they allow for the instantiation of abstract ideas in encompassing virtual worlds. This capability of the medium enables us to simulate reality in a cost-effective manner, for instance by treating anxiety through exposure therapy in virtual environments (Lindner et al., 2019) or training surgery on virtual patients (Satava, 1993). Beyond mere simulation, however, VR also holds the power to realize goals in ways that would otherwise not be possible, such as reducing implicit racial bias by embodying white people in dark-skinned avatars (Banakou et al., 2016) or increasing self-compassion by changing perspectives through virtual embodiment (Osimo et al., 2015). This latter approach—realizing goals in ways that would otherwise not be possible—involves the design of virtual worlds with ontologies different than the real world, tailored to elicit a particular effect on the immersed user. The power of VR to change ourselves in this manner is usually attributed to the capability of the medium to induce a feeling of presence in the computer-synthesized worlds (Slater and Sanchez-Vives, 2016). Immersed in VR, the user is *situated*; she feels present in the virtual environment, experiences it from a particular point of view, embodies avatars and tools, and involves herself in the scenario or narrative of the application. From this mediated situatedness, where some possibilities for experience are left open while others are restricted, a particular *subjectivity* of the user is constituted in relation to the *objectivity* of the virtual environment. Consider, for instance, how the embodiment of a child-sized avatar constitutes the virtual environment as large and perhaps overwhelming, or how the embodiment as a victim in a scenario may constitute the world as an unjust world in need of change. The user experience of VR is in this way dependent on how the subjectivity and objectivity of experience are constituted in *relation* to each other. What makes VR practical for interventions, of course, is that although the user's subjectivity is constituted in relation to a virtual environment, the effects are not restricted to the bounds of the simulation. The experience also plays a role in effectuating an altered human-world relation after exposure so that having experienced a virtual reality, reality itself is re-framed for the subject.

Because VR interventions owe their effectiveness to the experience of a virtually mediated subjectivity, we argue that insight into the phenomenology of these interventions can inform our understanding of them. In advocating for such a turn to experience, this paper presents a theoretical framework for understanding the user experience as *mediated* in relations constituted between user and environment. The mediation perspective that we advocate is distinguished from traditional approaches to understanding user experience in that it does not presuppose the human subject and the technology as poles between which interaction occurs (Verbeek, 2015a). Rather, it sees the human subject and the experienced technology as a *result* of this interaction as they "mutually shape each other in the relations that come about between them." (Verbeek, 2015a, p. 28). We purport this perspective is a more relevant way to understand the user experience underlying VR's capability to "change the self," as it specifically attends to how the human subject is mediated in the user-environment relation that is constituted.

A Postphenomenology of Virtual Reality

The framework we present for understanding and describing the virtually mediated experience is grounded in postphenomenology. Postphenomenology is a philosophy of technology that understands technologies in light of how they mediate human-world relations by co-constituting the subjectivity and objectivity of experience (Rosenberger and Verbeek, 2015). Postphenomenology is a highly relevant framework for understanding how VR technologies mediate experience, especially VR interventions, as these explicitly aim to change behavior, feelings, and attitudes, consequently, impacting the way that humans relate to their world. For instance, VR can be used to entice people to save for their retirement (Hershfield et al., 2011), enhance fear recognition in violent offenders (Seinfeld et al., 2018), or encourage prosocial behavior (Rosenberg et al., 2013). This is done by mediating a user-environment relation in VR within which the experience that effectuates the intervention takes place. Usually, this experience is approached in research through measuring several aspects of it such as presence, confirming the virtual embodiment, measuring simulator sickness, and generally accounting for a select number of psychometric variables. In this paper, we argue that approaching experience qualitatively from a broader post-phenomenological perspective can inform our understanding of the virtually mediated experience in a more holistic way than isolated constructs can offer. While a researcher studying user experience of VR from a post-phenomenological perspective would naturally also be concerned with whether a user feels present and embodied in the virtual environment, what she would have as her focus is how the embodiment and presence take part in constituting the user's subjectivity in relation to the objectivity of the environment. Approaching experience from a post-phenomenological perspective, therefore, does not involve replacing or rejecting established constructs used to measure experience; instead, it attends to this experience by describing it in terms of the subjectivity and objectivity arising from the mediation. For Immersive VR, this entails seeing the user experience as *mediated* in relations constituted between user and environment.

Ethics

Attending to the user experience of VR from a post-phenomenological perspective can also be useful for ethical assessment. The post-phenomenological approach to ethics is one of *ontological disclosure*; it asks what kind of worlds we disclose through new technologies, and in the same manner, who we become in relation to these worlds (Introna, 2017). Therefore, it is by providing an increased understanding of the ways that VR technologies can mediate our experience that the post-phenomenological perspective can aid researchers in discovering potential ethical issues resulting from their designs. Ethical concerns are particularly relevant for VR interventions as they explicitly aim to affect human behavior. We know that VR owes the effectiveness of its interventions to its mimesis of reality; the benefits observed in studies "rely on the extent to which the experience is perceived as real" (Slater et al., 2020, p. 1). In addition to the shared phenomenology of presence (Loomis, 2016), reality and virtuality also share what Metzinger (2018) refers to as *phenomenal transparency*, where the medium takes a transparent role so that the content it presents is not subjectively experienced *as* a representation. Consequently, it is because VR experiences can be similar to real life experiences (Slater, 2009) that VR is a powerful technology that is capable of producing beneficent as well as non-beneficent results. How complex the ethics of VR may become upon mass adoption is not known. Madary and Metzinger (2016) argue that VR will change deeply established notions of who we are and how we identify and so "transform the structure of our life-world" (p. 2). What is clear, however, is that the powerful capabilities of VR to "change the self" require researchers to exercise ethical attentiveness to the various ways in which a participant's subjectivity can change as the result of experiencing a virtually constituted subjectivity. Although the content of the experience is virtual, the experience is "*real as an experience*" (Slater et al., 2020, p. 5, emphasis in original), and the emotional and cognitive after-effects, although usually beneficial, can also be harmful (Slater et al., 2020). For instance, while VR interventions may reduce implicit racial bias (Banakou et al., 2016), they may also increase it in negative contexts (Groom et al., 2009; Banakou et al., 2020), suggesting potentially non-beneficent results when using VR as an "empathy machine." Similar warnings have been issued by Sri Kalyanaraman et al. (2010) who immersed participants in a simulation of the effects of schizophrenia. Although their simulation proved to be effective in increasing empathy and positive perceptions toward people who have schizophrenia in combination with non-VR perspective-taking exercises, they found that "mere exposure to a virtual simulation of schizophrenia by itself may not only be ineffective, but actually prove to be inimical…" (ibid, p. 441). Other non-beneficent results were also reported recently by Neyret et al. (2020) from a virtual recreation of a Milgram Obedience Scenario, who highlights it as "vitally important" to be aware of possible adverse outcomes resulting from virtual embodiment in scenarios—even if the chance of this occurring is deemed unlikely a priori.

Madary and Metzinger (2016) write how the embedding of VR in our world creates a "complex convolution, a nested form of information flow in which the biological mind and its technological niche influence each other in ways we are just beginning to understand" (p. 20). VR creates "not only novel psychological risks but also entirely new ethical and legal dimensions..." (ibid, p. 20). While no single approach or theoretical foundation can solve the ethical challenges of VR alone, we believe a qualitative turn to the user experience of VR—by inquiring into the experiential relationship established between user and environment—can be a complementary constructive angle from which researchers can uncover unintended effects resulting from their designs.

This paper is structured as follows. First, we provide a background to postphenomenology and account for its relevance as a framework for describing Immersive VR mediation. Having presented the paper's theoretical background, we detail our proposal of a post-phenomenological framework for understanding user experience in Immersive VR as mediated in *user-environment relations*. We demonstrate the applicability of the framework by analyzing a selected variety of studies on VR interventions that constitute particular user-environment relations in order to mediate particular human-world relations. After the analysis, we discuss the relationship between real and simulated subjectivity as well as the relationship between real and virtual worlds in more depth. Finally, we discuss the scope of the framework before outlining directions for future work to advance the applicability of the theoretical framework into the methodological.

RELATED WORK

Postphenomenology

The framework that we propose in this paper is informed by postphenomenology, a philosophy of technology that views technologies as mediators of human-world relations. With its phenomenological roots, postphenomenology understands humans and technologies as inseparable and views technologies as co-constituting human subjectivity and world objectivity (Rosenberger and Verbeek, 2015). Consider, for instance, how the embodiment of a car enhances the human being by constituting the subject as a driver and therefore also the world as more accessible or how, for a blind person, the white cane constitutes the world as such and extends the subject through the embodiment of the cane. Concerned with empirical data (Achterhuis, 2001), postphenomenology is pragmatic, and giving heed to its phenomenological origins, it draws its data from experience. Postphenomenology adopts from phenomenology the notion of *intentionality* as an invariant of experience: all consciousness is consciousness of something. Subjectivity and objectivity, experiencer and experienced—what Husserl referred to as the *noesis* and the *noema*—are two distinct ends of the polarity of experience. Postphenomenology stresses the role that technologies have in mediating this intentional relation by co-constituting both the human subject and their world. In doing a post-phenomenological investigation of a VR application, therefore, we would be interested in "who" the user becomes in relation to the virtual environment, and simultaneously, "what" the environment is for the user. In other words, we would be interested in what kind of user-environment relation is being mediated, but also beyond this, how the user-environment relation takes part in mediating the human-world relation outside of the virtually mediated experience.

Postphenomenology as a praxis-oriented phenomenology was established through the works of philosopher Don Ihde. An expanding group of scholars now contribute to the post-phenomenological approach of studying the ever-expanding role of technologies in our lives, most notably Peter-Paul Verbeek, who extends Ihde's post-phenomenological thought in his theory of technological mediation (Verbeek, 2005a). In the sections below, we provide an account of Ihde's Human-Technology Relations before describing Verbeek's exposition of immersion as a human-technology relation.

Human-Technology-World Relations

Don Ihde identified four structures of human-technology-world relationships (Ihde, 1990). The first of these he calls *embodiment* relations, where the combination of human and technology together relate to the world. In embodiment relations, there is transparency, as when we look *through* our eyeglasses or talk *through* the phone. Second, he discusses *hermeneutic* relations, where humans "read off" an abstract representation by a computer, such as a weather forecast or an MRI scan. Third, in *alterity* relations, humans interact with technology directly within its own system, a common example being interaction with an ATM or a calculator, where the world withdraws into the background. Lastly, Ihde (1990) discusses what he calls *background* relations, where the technology is an implicit condition affecting the environment, partly serving as the context in which we find ourselves (e.g., an air conditioner). Ihde (1990) illustrates his embodiment, hermeneutic, alterity, and background relations through diagrams indicating on

TABLE 1 | Human-Technology Relations Diagram (Ihde, 1990).

Embodiment relation	(human – technology) → world
Hermeneutic relation	human → (technology – world)
Alterity relation	human → technology (world)
Background relation	human → (technology / world)

Arrows indicate intentionality.

which poles, subjective or objective, the technology primarily is "situated" with arrows indicating intentionality, as seen in **Table 1**.

Immersion as Human-Technology Relation

The human-technology-world relations identified by Ihde are not so exhaustive as to include all possible relations. Verbeek has further identified several human-technology relations enabled by newer technology developments, where the immersion relation is the most relevant for the user experience of VR. The immersion relation can be understood as a more active version of Ihde's background relation, where the environment and the technology become merged (Verbeek, 2015b; Aydin et al., 2019). It is more active in the sense that the environment is aware of human beings and actively interacts with them. The result is that human beings are directed toward technologies, and the technologies are in turn directed toward them, resulting in a "reflexive intentionality" (Verbeek, 2005b) where humans can have new relations toward themselves through the technology. Although this relation is referred to as an "immersion" relation, we should note that Verbeek does not use the word "immersion" in order to relate it to VR technologies in particular. As examples of immersion relations, Verbeek (2011) describes smart toilets that analyze excrement and provide health reports, or beds that can detect whether somebody falls out. The immersion relation is nevertheless relevant for understanding VR because VR technologies open entirely new possibilities for reflexive intentionalities, which we return to in our analysis.

User-Environment Relations

Having described Ihde's and Verbeek's human-technology relations, we might ask what kind of relation VR constitutes. As we have discussed, the benefit of VR is its flexibility; it can be adapted to unique situations and be designed to elicit vastly different effects. In this regard, VR can be said to be an extreme meta-medium (Kay and Goldberg, 1977), as virtually all other media can be reproduced within it, including future, non-existing media. The result is that "…each form of VR is a medium unto itself." (Lanier, 2017, p. 204). For this very reason, any attempt to give a total account of the various possibilities of VR mediation is impossible; all the various human-technology relations introduced above could conceivably be had *within* various VR applications. There is an invariant human-technology relation that lays the ground for other relations within the virtual, however, it takes a special form in VR. Comparing immersive VR to non-immersive simulators, Voordijk and Vahdatikhaki (2020) write that "when the technology 'disappears' in embodiment, the role of the VR simulator changes, in terms of Ihde, from an alterity relationship to an embodiment relationship." (p. 10). While the VR HMD becomes transparent in use and we act through it, the intentional relation is not mediated toward the world, rather, it is mediated toward the virtual environment. Consequently, when embodied, the user is in an alterity relation toward the virtual environment, interacting directly with the technology within its own system. Thus, in the embodiment of a VR HMD, we act both through it and upon it, which is why VR can simultaneously mediate both (i) the objectivity of the environment in which users are situated (alterity) as well as (ii) the users' subjective position and relation toward that virtual environment (embodiment). So, while we embody parts of the VR technology (hardware, avatars, tools) as part of our subjectivity in a transparent embodiment relation, the objectivity of our experience (environment, actors, social scenarios) is also mediated by the same VR technology, constituting an opaque alterity relation in which the world is in the background. This human-technology relation that VR constitutes, we describe in our framework as *user-environment relations.* This embodiment-alterity relation can be schematized in the manner of Ihde (see **Table 1**) as follows:

$$(\boldsymbol{Human} - \boldsymbol{Technology}) \rightarrow \boldsymbol{Technology}(-\boldsymbol{World})$$

This schema denotes a *user-environment relation*: a human in an embodiment relation with the technology (i.e., the user) in an alterity relation to the technology (i.e., the environment), while the world is in the background.

A POST-PHENOMENOLOGICAL FRAMEWORK FOR IMMERSIVE VIRTUAL REALITY

Immersive VR mediates user-environment relations in which the embodied user stands in an intentional relation to the environment while the world is in the background. This human-technology relation that VR constitutes lays the ground for our framework of VR mediation. In substantiating our framework, this section will present and discuss the constitutive elements of this mediation process in more depth. As illustrated in **Figure 1** [which is an altered version of Figure 3 by Hauser et al. (2018) depicting the roles of design researchers in RtD inquiries] this means recognizing the subjectivity-objectivity structure as constituted within VR (the user-environment relation), as well as the subjectivity-objectivity structure as constituted outside of VR (the human-world relation).

Our framework mirrors the overview of technological mediation provided by Hauser et al. (2018); the humans of the study, the mediator, and their world are the basic constituents of any technological mediation process (see **Figure 2**). As the technology mediates the humans' subjective relation to their worlds, who these people are, where they are situated, and what the technology/mediator is are essential overarching variables in understanding technological mediation post-phenomenologically (Hauser et al., 2018). In post-phenomenological inquiries in Human-Computer Interaction, the researchers stand in constructive roles regarding the studying,

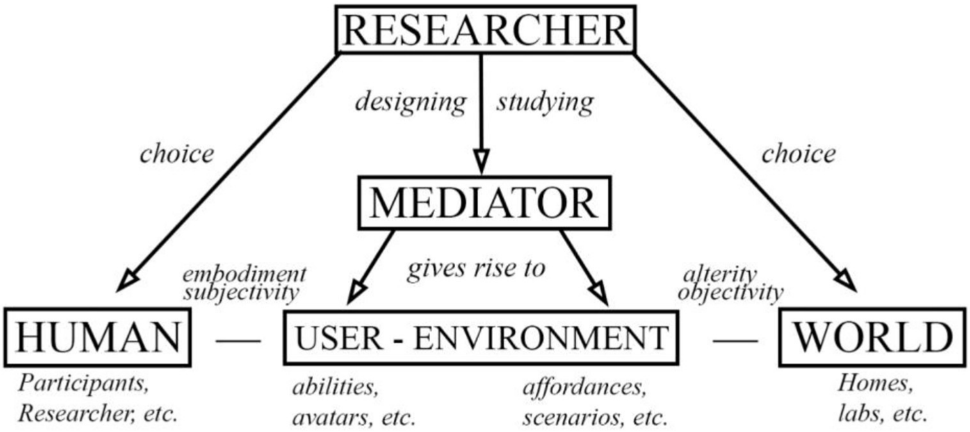

FIGURE 1 | Illustration of our Post-phenomenological Framework for Studying User Experience of Immersive VR as Mediations.

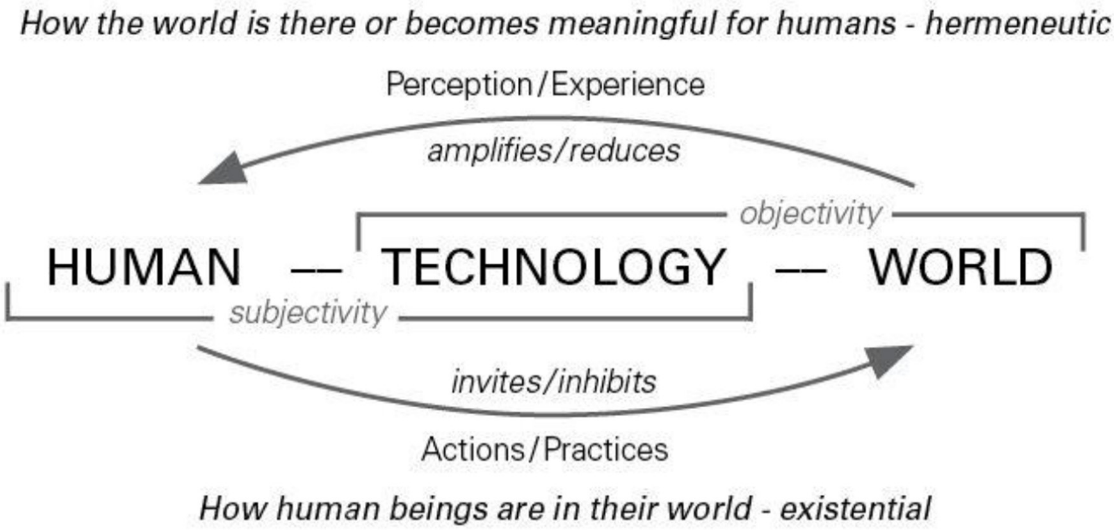

FIGURE 2 | Technological mediation (Hauser et al., 2018).

choosing, and designing of these constituents (see **Figure 3**). The next two elements in our framework more concretely address the user-environment relation: what occurs when a human participant engages with the VR application. The VR is here a mediator that gives rise to (4) a *User,* and (5) an *Environment*; the human as user has an altered subjectivity constituted in relation to the virtual environment.

In the next sections, we detail the various elements of our framework. An overview of the framework components is provided in **Table 2**.

Human

The first element of the framework is concerned with the human who engages with the VR mediator. Postphenomenology sees technologies as multi-stable, the same technology can have several different *stabilities* in terms of how it is used and experienced. While multistability of technology can be actively designed for—our best example being the smartphone, the modern swiss army knife—multistability is also present in cases where the intention is for the artifact to embody a concrete function, such as a VR application intended to deliver a particular intervention effect. In short, technologies "simply can't be reduced to designed functions" (Ihde, 2002, p. 106). As a classical example, hammers are made for hammering nails, but can find other stabilities, such as being a paperweight or a weapon (Ihde, 2002). In the same way, an interactive VR application is not fixed in how it can be "used" or experienced, the user-environment relation that is mediated depend not only on the VR application, but on the individual human who engages with it in their context of use. The particularities of this group, such as their sedimented or unestablished relationship with VR technology, or their attitudes toward technology in general, will impact their virtually

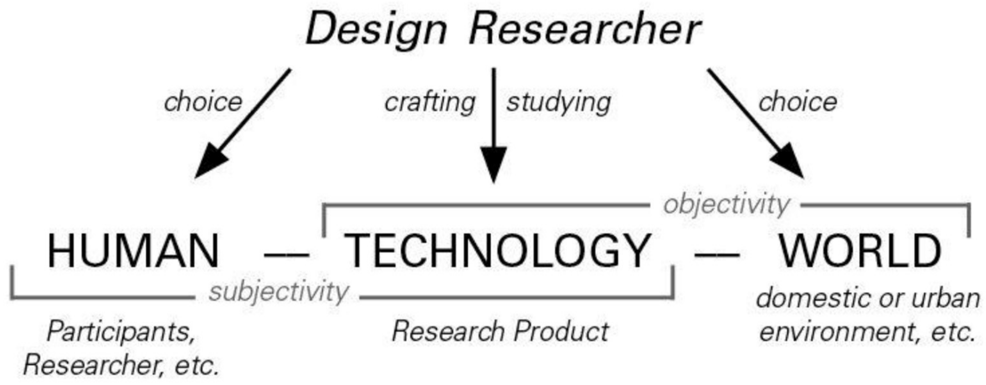

FIGURE 3 | Roles of design researchers in their RtD inquiries (Hauser et al., 2018).

TABLE 2 | Overview of Framework Constituents.

1. Human. *The human being partaking in the study. Here, the particularities of the person might be mediated, as well as impact the mediation. Examples of human factors: personality, gender, socio-economic status, interests and motivations, involvements, and previous technology experience. Human factors vary and impact relational and hermeneutic strategies toward the technologies*
2. World. *The use context of the application where the VR application is being used. This constitutes the background of the VR experience. Here, also, the particularities of the context might be mediated, or take part in mediating the VR experience. Examples: hospital, lab, work, or domestic settings*
3. VR Mediator. *The VR application that is being designed or evaluated for intervention purposes. Designed or studied for its ability to provide an experience or user-environment relation that can be a catalyst for change*
***4.* User.** *The human as user in an embodiment relation to the alterity of the virtual environment. The user subjectivity is in a nested relation to the subjectivity the human individual has in relation to her actual world (Gualeni and Vella, 2020) but is further affected by avatars, tools, interaction possibilities, position, involvements, and social scenarios*
5. Environment: *The virtual environment as experienced by the user during the VR embodiment. The part of the VR application that is not embodied by the user, but is rather acted upon, or that which acts upon the user, including social actors, 3D objects, events, etc.*

mediated experience. These are the humans who will experience the mediating effects of the technology on their self as well as on their world. Professional skills or diagnostic criteria fall under this point, but also differences in experience, culture, gender, etc., as this may be indicative of different *relational strategies* in approaching the technologies. Relational strategies can be defined as particular understandings and bodily approaches that "enables a user to relate to a technology in terms of a particular stability" (Rosenberger and Verbeek, 2015, p. 29). Similarly, different people may have different *hermeneutic strategies*, strategies which "enable a user to apprehend the meaning of a technology's readout in terms of a particular stability..." (Rosenberger and Verbeek, 2015, p. 29). This does not mean that VR applications can be so multi-stable that any user-environment relation can be experienced; as Ihde (2002) notes, "Multistability is not the same as neutrality." (p. 106). All technologies, however open they may be, have a certain *directedness* (Verbeek, 2008). Despite there being various trajectories for use, these are not unlimited, and some of these will prove more dominant and stable than others.

World

The second element of the framework is concerned with where in the human sphere the mediating technology is used. Phenomenological accounts of places and situational contexts highlight the inseparability of humans from their environments (Donohoe, 2017). Places—understood geographically, architecturally, or socio-culturally—take part in shaping behavior, identity, and moods; places, then, can also be regarded as mediators of our selves. The humans in the study, therefore, are only one aspect; equally relevant is the "world" in which these humans will use the technology. The world is the "use-context" of the application and will be present for the user as the background of experience, although she is immersed in a virtual world. Again, we return to the post-phenomenological concept of multi-stability; technologies will have different meanings for different people *in different contexts*. While it is possible to evaluate an application experimentally in-lab that is originally intended to be used elsewhere, this is not as likely to give an account of what the mediation effects of the technology will in fact be, simply because one of the constituents of the virtually mediated experience (the world) will be different than what is being evaluated. This is equally due to the situational context (being observed by researchers in-lab) as the geographical context of being in the lab. The use-context affects the experience of the virtual environment and the virtual environment may also further mediate how the participant sees the use-context.

Mediator

The third element in the framework is the technology, or mediator, which within VR interventions usually comprises a virtual environment that is experienced from the first-person point of view. The design of this technology can have various goals—such as therapy or training—which is meant to be attained by immersing the human into the virtual environment. This mediator gives rise to a user-environment relation: the human becomes an embodied user, immersed in, and standing in an intentional relation to, the virtual environment, while the world is in the background. In order to describe the mediator element more thoroughly, we detail the mediations that it gives rise to in the next sections: (4) User and (5) Environment.

User

The first sub-element of the mediator is concerned with the embodied user, situated and operating from a particular subjective position within the simulation. The "user" entity is not the same as the human participant, nor is it merely the subjective position into which the participant is immersed. Rather, it is the human participant *as* user, i.e., the human participant under active mediation of the VR technology, virtually embodied and in an intentional relation to the virtual environment. The subjectivity of the user, therefore, can be said to be "in a nested relation to the individual's subjectivity in the actual world" (Gualeni and Vella, 2020, p. xxi). Human subjectivity is being mediated by the VR application, within the simulation in relation to the virtual world (User) as well as outside the simulation in relation to the real world (Human). This is an example of what De Mul (2010) refers to as *poly(ec)centric positionality*, denoting a mediated multiplication of one's center of experience. Phenomenologists such as Merleau-Ponty (2002) have famously distinguished between the body as *lived* and the body as *object*. This holds for VR also; while our avatar can be seen as one 3D object situated amongst others, it is also, to the degree that it is embodied, that through which we experience. Ihde (2002) refers to Merleau-Ponty's lived body as *body one* and understands the objectified *body two* as the acted-upon body of Foucalt; "…body one permeated with the cultural significances that are also experienced." (p. xviii). According to De Mul (2010), it is this eccentricity—our being simultaneously inside (subjectivity) and outside (objectivity) of ourselves—which is the condition for telepresence and VR. With these technologies, we can objectify our thoughts of who we want to be, and, through embodiment, we can experience reality from the perspective of these bodies. In poly(ec)centric positionality, the virtual constitutes "a complete and additional, artificial experiential center" (Gualeni, 2015, p. 115) which lays the foundation for the simulation's capability to "elicit ontological effects" (ibid, p. 118).

The question of the user element in the framework is how this new, artificial experiential center is experienced during embodiment. It is concerned with what kind of subjectivity is mediated within the user-environment relation. For instance, what avatars and tools does the user embody? How is she positioned in relation to the virtual environment, and what are the possible points of action from this situatedness? Is she involved in a certain story, scenario, or task? Here, examples may range from leading a team of surgeons, to being positioned as a victim of physical abuse. What the user can do, and who she experiences herself to be, is defined in relation to the virtual environment and the affordances it presents.

Environment

In strong relation to the user, therefore, is the environment, the second sub-element of the mediator. The environment represents the part of the VR application that is not embodied, and therefore, that to which the embodied user relates to as alterity. In focus is the question: in what kind of environment or world is the user situated? What are the basic parameters for how this environment works and what it represents? For instance, the system may display some objects as interactable and some merely acting as decorative or situational elements, some in the proximity of the user and some at a distance. Such choices are a part of the intended mediation of the researcher, impacts the subjectivity of the user, and provides the technology with a certain *directedness* (Verbeek, 2008). It is here important to note that we understand the environment (post)phenomenologically. We are interested in how the environment is understood from the situated standpoint of the user, not from a detached God's eye view. Similar questions exist here as for the world constituent, but in relation to the environment. For instance, *as what* is this virtual environment disclosed for the subject? What are the most apparent features or affordances of this environment, and what does this communicate to the user? Does the environment invite certain trajectories of action, while inhibiting others? In other words, we are interested in how the virtual environment is experienced in its relation to the user, that is, how the human perceives the environment when immersed and embodied. How the environment is experienced is not just dependent on the objective features of the environment. An illustrative example of such numerous convoluting, mediating factors is the various virtually reconstructed Milgram Obedience Scenarios (Slater et al., 2006; Gonzalez-Franco et al., 2018; Neyret et al., 2020). In these virtual recreations of the famously controversial research described by Milgram (1964), participants meet with real researchers in a lab who do an experiment, but the experiment is to meet virtual researchers and partake in their experiment in a virtual lab. In the event of partaking in such a study, participants are in a very real sense both real and virtual participants, and likewise, stand in relation to both the real researchers and the virtual researchers. VR technologies are not immersive to the extent of the participants forgetting their normal feeling of self or their worlds. The real world is still present as a background relation, and the user subjectivity is in a nested relation to the subjectivity of the human individual in the real experiment.

ANALYSIS OF USER-ENVIRONMENT RELATIONS

VR constitutes an embodiment-alterity relation that we describe as user-environment relations, where the embodied participant is in an intentional relation toward the alterity of the VR application. This describes VR mediation generally; how a *particular* user-environment relation is mediated depends on *what* is embodied (subjectivity), and *what* is related to as alterity (objectivity). In order to concretize our framework, this section presents an analysis of various user-environment relations constituted in VR interventions. What we intend with our analysis of user-environment relations in VR is to account for some observed variance of how user-environment relations can

TABLE 3 | Overview of User-Environment Relations from Analysis.

Simulated Subjectivity: *Simulated Subjectivity refers to mediations where an altered subjectivity is experientially pronounced; a simulation of "what it is like." This may be done with the intent of having the application act as an empathy machine or for the application to facilitate for the experiencing of new first-hand perspectives on known information. Examples include what it is like to suffer from strokes, schizophrenia, blindness, as well as the effects of psychedelic drugs*
Simulated Objectivity: *Simulated Subjectivity refers to mediations where the user is "transported" to a new place, where there is little to no explicit attempt to alter user subjectivity apart from being immersed in the environment. Examples include medical simulations, cultural heritage, VR exposure therapy, etc.*
Subjectivity-Objectivity Inversion: *Subjectivity-Objectivity inversion refers to a mediation process in which the subjectivity-objectivity poles are inversed, for instance framing the Self as Other, or the Other as Self to change either one's perspective on one's self or one's perspective on others. What "other" is being re-framed (opposite gender or different age/race/socio-economic status) varies depending on the Human-World Relation the researchers intend to achieve (for instance increased empathy or less racial bias)*
Subjectivity-Objectivity Synchronization: *Subjectivity-Objectivity Synchronization refers to a mediation process in which the subjectivity and objectivity of experience approximate each other toward a state of equilibrium. This can be initiated by mediating properties of the subjectivity to affect the objectivity or the other way around. Which mirrors which can depend on what Human-World Relation the researchers intend to achieve or measure*

be structured. The research papers in the analysis were selected in order to display the breadth of ontological structuring that is possible within the overarching embodiment-alterity relation. The analysis highlights in post-phenomenological terms how the interventions constitute various user-environment relations in order to mediate various human-world relations.

We categorize the identified user-environment relations as follows: (1) Simulated Subjectivity, (2) Simulated Objectivity, (3) Subjectivity-Objectivity Inversion, and (4) Subjectivity-Objectivity Synchronization. The first two categories focus on the two distinct poles of experience in VR: subjectivity (embodiment) and objectivity (alterity). These are discussed rather briefly, and by dealing with subjectivity and objectivity in isolation, these categories also act as an introduction for the two latter categories where subjectivity and objectivity are more entwined. Consequently, the analysis is mainly concerned with the two latter categories, "Subjectivity-Objectivity Inversion" and "Subjectivity-Objectivity Synchronization", as these describe the novel *relations* that can be constituted between user and environment in VR. A summary of the identified user-environment relations is provided in **Table 3**.

Simulated Subjectivity

In providing an experience, VR mediates sensory stimuli, some of which is embodied and becomes "part of" the user, and some of which is not embodied, and as such stands in an alterity relation toward the user as an environment. While this means that all VR applications will necessarily simulate both subjectivity and objectivity, what is novel or unique in the VR experience may be more pronounced experientially for the user. Simulated subjectivity, therefore, refers to cases where the intended mediation is to convey *what it is like* to be another (subjectivity), with less focus on mediating a particular virtual environment (objectivity). It refers to cases where it is intended for the mediation of subjectivity to be more pronounced experientially than the objectivity. As an example, Suzuki et al. (2017) developed the *Hallucination Machine* by processing panoramic videos using Google's Deep Dream AI, in order to "[induce] visual phenomenology qualitatively similar to classical psychedelics." (p. 1). Other examples of simulated subjectivity include simulations of various visual impairments in VR (Ahn et al., 2013; Ates et al., 2015; Jones et al., 2020) as well as strokes (Maxhall et al., 2004) and schizophrenic episodes (Nyre and Vindenes, 2020). Simulating subjectivity is naturally linked to empathy as it could be said to be a virtual representation of "what it's like to walk a mile in someone else's shoes." However, most of the interventions promoting empathy in our analysis are discussed under section Subjectivity-Objectivity Inversion, as their strategy toward generating empathy is by mediating a more reflexive user-environment relation in which the alterity/objectivity is also of importance.

Simulated Objectivity

As the inverse of Simulated Subjectivity, Simulated Objectivity refers to mediations when the participant is immersed in an environment or scenario (objectivity) where there is no explicit intention of altering user subjectivity. Typical examples here include simulator training for various purposes such as surgery (Alaraj et al., 2011), but can also be exemplified through virtual field trips (Çaliskan, 2011), cultural heritage (Rua and Alvito, 2011), or VR exposure therapy (Flobak et al., 2019). In these cases, the success of the simulation is dependent on the degree to which the simulation represents reality. This is VR as it perhaps is traditionally understood, where the participant is "transported" to an environment but remains "herself." Thus, there is the intention of keeping the participant's subjectivity more or less non-mediated, apart, of course, from the mediating effects of the environment/situation itself.

Having briefly described Simulated Subjectivity and Simulated Objectivity as the two distinct poles that can be targeted in mediation, we move on to the reflexive user-environment relations, where the structured relationship between subjectivity and objectivity is of importance. Naturally, the two next user-environment relations also include the simulation of subjectivity and objectivity, but here it is the user-environment *relations* that are highlighted.

Subjectivity-Objectivity Inversion

In this section, we present a user-environment relation that we refer to as a *subjectivity-objectivity inversion*. We discuss this from two angles: mediating the Other as Self and mediating the Self as Other.

Other as Self

As humans, we identify in particular ways. We identify as individuals, but also with particular groups, such as socio-cultural, racial, and ethnic groups, as well as gender and age. To various extents, other groups are experienced as such, *other*, and so we experience ourselves and our own situation in a different perspectival manner than we do others and their situations.

While this is a natural limitation of being a particular human being, VR can allow a user-environment relation that constitutes what has traditionally been related to as Other (objectivity) as Self (subjectivity). The studies which we cite below as examples of this usually comprise an active instantiation of perspective-taking (van Loon et al., 2018) where VR allows the point of perspective to be an actual experiential center as opposed to one imagined through cognitive activity.

An example of such a subjectivity-objectivity inversion, *Other as Self*, is present in the study by Banakou et al. (2016), who embodied 90 white females in black virtual bodies. They found an immediate decrease in implicit racial bias against black people. A similar experiment was performed by Hasler et al. (2017) who embodied 32 white females and 32 black females in avatars of various color so that, over two sessions, all participants had been embodied in both black and white avatars. They found that the embodiment enhanced mimicry of behavior between those of the same embodied racial group—independently of the actual race. Similar role changing by means of virtual embodiment has been conducted by Seinfeld et al. (2018) who embodied male domestic violence offenders in virtual female bodies where they experienced a virtual scene of abuse from a first-person victim perspective. After exposure, the male offenders had an improved ability to recognize fear in female faces, a trait which offenders as a group score significantly lower on compared to controls (Seinfeld et al., 2018). Other examples include embodying adults as children (Tajadura-Jiménez et al., 2017; Hamilton-Giachritsis et al., 2018), embodying younger people as elderly (Hershfield et al., 2011; Banakou et al., 2018), or even embodying animals (Ahn et al., 2016). In all these cases there is a perspective-taking where what was traditionally conceived of as outside one's subjectivity enters within it. What this "other" should be depends on the kind of intervention that is intended. Mothers may get an increased understanding of what it means to be a child, which in turn may alter how they view their role as mothers. Younger people may experience what it is like to inhabit an aged body, perhaps altering how they view the impermanence of their youth and the role of their elders, and people embodied as animals may feel more connected to nature by being directed to reflect on the fact that animals are sentient too. By reframing what is mediated as the subjectivity and objectivity of experience, VR can through subjectivity-objectivity inversion help humans bypass sedimented relations and facilitate a perspective-taking that is more directly experienced.

Self as Other

Another example of subjectivity-objectivity inversion is the reframing of the *Self as Other*. Just as being a particular human being comes with a limited perspective of others, seeing ourselves from our own point of view can have its limitations as well. "From the perspective of the self, the other is so rounded out that it is a consummated, self-sufficient whole. In contrast, the self cannot see itself in that way. It is tied up in the incompleteness of its own story…" (McCarthy and Wright, 2004, p. 75). While we may be able to see others for who they are now, we see ourselves in terms of both our future and our past. Being caught up in worries for the future and regrets from the past may cloud our access to the present reality. Objectifying the self, therefore, may come with its own benefits of altered perspectives. The studies cited below usually comprise a more active instantiation of self-distancing theory (Leitner et al., 2017) of which methods are traditionally performed through the imagination. An example here is the study presented by Osimo et al. (2015), who had male participants embodied in avatars closely resembling themselves describe a personal problem to a virtual person in the likeness of Dr. Sigmund Freud. When the participant has described his problem, his body is swapped to that of Freud's, now seeing the avatar created in his likeness, which he previously identified with, sitting opposite him. Then his avatar begins to tell the story he had just told back again to the user embodied as Freud. Here, the participant as Freud again answers in terms of advice, before swapping back to the avatar again, and so on. In this way, the application reframes the self as other, as well as the other as self, and ideally allows the user to address his own problems as he addresses others' problems. Osimo et al. (2015) write how "…this form of embodied perspective-taking can lead to sufficient detachment from habitual ways of thinking about personal problems, so as to improve the outcome, and demonstrates the power of virtual body ownership to affect cognitive changes" (p. 1). A study similar in mechanism was conducted by Falconer et al. (2014) where female participants were trained in providing a compassionate response, which they delivered to a child in VR while embodied in a (non-lookalike) adult body. Later, the participants experienced their own compassionate statements in the embodiment of a child, which the researchers found increased self-compassion and feelings of being safe. Here, the perspective-taking which the body-swapping facilitated (i.e., the alteration of subjective roles) allowed the participants to be both on the giving and receiving end of compassion. Another example is brought forward by Bourdin et al. (2017) who created out-of-body experiences in VR by embodying participants in avatars, and changing the viewpoint so that they could view their virtual bodies from outside, reducing fear of death in the participants. Our final example of a subjectivity-objectivity inversion is the embodiment of participants as older versions of themselves in order to promote saving for their retirement (Hershfield et al., 2011). Here, the participants embody their future selves as part of their subjectivity and look in a virtual mirror. What is "other" in this intervention, however, and which the researchers intended the participants to identify more strongly with, is the aging of this future self. This can also be done where the "other" is not age deterioration, but increased/decreased physical fitness in order to increase motivation (Fox and Bailenson, 2009).

In the user-environment relation we call subjectivity-objectivity inversion—self as other and other as self—what the human participant embodied as user relates to as themselves is inversed. The result is that what was previously embodied (subjectivity) is now the alterity (objectivity), or that what was previously alterity is now embodied. This makes for an immersion relation between the user and the environment which constitutes a reflexive intentionality where the user can experience standing in new relations to themselves and others. We reiterate that reflexive intentionalities occur when the human is in an intentional relation to the technology-infused environment, where the technology-infused environment is also directed in intentionality toward the human. The human can experience how the environment perceives or interprets her from its perspective. In VR, however, the reflexive intentionality is realized somewhat differently. Firstly, the technology is fused with the environment in the sense that the technology is what instantiates the environment as such. Further, the environment does not abstract or convey a "representation" to the user of how it perceives her, which the user is meant to see from her situated perspective. Instead, aspects of the virtual environment that the user stands in an intentional relation to, such as a social actor, can itself be embodied so that the new relation that is opened toward one's self can be experienced more directly.

Subjectivity-Objectivity Synchronization

Having described Subjectivity-Objectivity Inversion, we turn to the case of Subjectivity-Objectivity Synchronization. A subjectivity-objectivity synchronization is an attempt at producing harmony between the inner life of the user and the external world that is experienced. The attempt can either be to make the inner life of the user be represented through the external world, or to make the external world affect the inner life of the user, or both. In the way that subjectivity-objectivity *inversion* utilizes an active instantiation of perspective-taking and/or self-distancing techniques, applications facilitating subjectivity-objectivity *synchronization* actively instantiate meditative techniques such as Mindfulness. Many meditation or relaxation techniques have as their aim to redirect focus and attention on the breath or the body in order to promote a feeling of union both with oneself and the world. In VR, the attempt to promote unity between subjectivity and objectivity—self and other—is approached explicitly by blurring distinctions or creating new relationships between the two. For instance, Roo et al. (2017) created a mixed reality sandbox where the user can create a virtual environment by restructuring sand in a physical sandbox. The sandbox has an overhanging depth sensor measuring the peaks and valleys of the sandbox, and a projector that projects visual terrain upon it. Having created the environment, the user can immerse herself in a 3D render of this world through an HMD where the environment responds to physiological data of the user, such as breath and heart rate. Here, the aim is to facilitate mindfulness meditation through a focus on the body as it is mediated through the environment. The mediation amplifies the focus on bodily sensations such as breath and heart rate, and by having this represented in the external environment, the otherwise clear-cut boundary between self and other is diminished so that there is subjectivity in objectivity and vice versa. A similar example is brought by Amores et al. (2019) who designed "Deep Reality," a VR experience of underwater fluorescent beings that move based on biometric information such as electroencephalogram (EEG), heart rate (HR), and electrodermal activity (EDA). The aim was reflection and relaxation. Here, again, the recurring pattern is that of changing the external environment to affect inner states, and as with Roo et al. (2017), the external environment is in turn based on inner states or approximations of these, constituting a neurofeedback loop in which it is intended that the subjectivity and objectivity of experience should approximate each other toward a state of equilibrium. Another example is brought forth by Stepanova et al. (2020) who designed JeL, an immersive VR system designed "to bring awareness to our physiological rhythm, fostering a connection with our bodies, each other, and nature (p. 641). Here, two users aim to synchronize their breath in order to grow corals in a coral reef. Other examples include the projection of artistic visualizations in VR based on EEG in order to induce positive pre-sleep (Semertzidis et al., 2019), biofeedback through projection to support yoga-breathing practices (Moran et al., 2016), and virtual environments generated by users' brain activities and respiratory rates in order to assist novice users in learning to reduce stress through mindfulness mediation (Prpa et al., 2016).

In these user-environment relations, the users also stand in an intentional relation toward the environment and so experience the environment, and likewise, the environment is in an intentional relation toward the user and "experiences" the user. In the study by Semertzidis et al. (2019), for instance, where the EEG is artistically visualized, the user perceives how the mediator interprets her state. This makes for an immersion relation between the user and the environment and opens up for a reflexive intentionality where the user not only experiences the environment, but a new perspective is opened toward one's self. Depending on the extent to which the user attempts to read or interpret the "message" of the application, these relations may lean toward hermeneutic as opposed to alterity.

This concludes our analysis of user-environment relations in VR interventions. We wish to stress that this list is far from exhaustive, and that the user-environment relations do not necessarily exclude each other. It is perfectly possible to imagine combinations of these as well as other possible subjectivity-objectivity configurations. We return to the idea of VR as an extreme meta-medium: each VR application constitutes its own form of medium. Beyond what we have described above, every user-environment relation will have its own subtly differently constituted subjectivity-objectivity structure, and we expect more nuances and complexity as researchers relate to actual phenomenological accounts. As Ihde (2012) writes regarding the methodology of phenomenological investigations, "[t]he analysis begins with *what* appears (noema) and then moves *reflexively* toward its *how* of appearing [noesis]" (p. 31). What kind of subjectivities will be revealed in virtual worlds cannot be grasped beforehand; this is rather discovered reflexively based on the mediated experience.

DISCUSSION

Interaction with technology is traditionally understood as something that happens between the human being and the technological artifact (Verbeek, 2015a). In contrast, postphenomenology takes the perspective of understanding the human subject and the technological artifact phenomenologically as they arise from the interaction; it pays attention to how the human subject and technological artifact mutually shape each other in the relation that comes between them. The perspective sees the design of technological objects as also involving "the design of human subjects who interact with these objects." (Verbeek, 2015a, p. 28), making it particularly relevant for understanding the user experience of VR interventions whose aim it is to "change the self." The theoretical framework in this paper is proposed as relevant for describing both intended and actual VR mediations. In order to clarify the contribution of the framework, we discuss more in depth the relationship between real and "virtual" subjectivity, as well as real and virtual worlds, before discussing the scope of the framework. We end the discussion by outlining directions for future work of advancing the applicability of our framework into the methodological.

The Relationship Between Human and User

Attempting to understand the nuances of the fleeting and mediated experience of VR can be complex. VR is a personal experience and will alter (and depend) on who the participant is, and in which world of meaning that they live. While the VR application is constant, the lived VR experience is a transaction between the technology and the human. So how exactly is this relationship constituted? To draw an example from post-phenomenological literature, Kaposy (2017) looked at how simulating ethical scenarios in medical education purports a view of the medical student more as an object than a subject. Utilizing Ihde's distinction of body one and body two—body one being the subjective, lived body, and body two, the objectified social and cultural body—the insight by Kaposy (2017) is that the students within the scenario are being evaluated after certain objective criteria, constituting an expected way of being that is abstracted as an object body. This is also the nature of interventions in Immersive VR. Within the design, the role that is more or less adopted upon embodiment and defined in relationship to social actors and the virtual environment is an abstract object body, a "body two." We draw on information from our environment and our bodies' appearance in determining who we are, and this impacts our behavior. This is, of course, not just a phenomenological discovery. This nested subjectivity is also described within other disciplines. For instance, both The Proteus Effect (Fox et al., 2013) and the idea of Body Semantics (Slater and Sanchez-Vives, 2014) claim, and demonstrate, that body type can influence attitudes and behaviors. The Proteus Effect describes the mechanism utilized in many VR interventions from a social psychological perspective based on self-perception theory, where participants conform to the behavior they imagine that a third party would expect (Slater and Sanchez-Vives, 2014). Body semantics approaches this from a neuroscientific perspective and sees this as an intrinsic property of brain functioning, where the brain generates attitudes and behaviors "concomitant with that type of body, independently of any other factors such as social expectation." (Slater and Sanchez-Vives, 2014, p. 28). Returning to the example brought forth by Kaposy (2017), however, the point is that although we may embody an objectified "body two", it does not fully become who we are. Kaposy (2017) underscores the need to recognize the "anthropological constant" of bodily lived experience (body one) in the simulated clinical encounter. Although body one will never 'become' body two after long enough exposure, there is here a synthesis: "body one is situated within and permeated with body two, the cultural significations which we all experience." (Ihde, 2003, p. 13). Consequently, in VR, our "virtual selves" and virtual worlds—and how they are ontologically structured—do not become our new selves and our new worlds. They do, however, affect the way the "real world" and our "real selves" are constituted. Take for instance the study by Banakou et al. (2016), in which white participants were embodied in black avatars. The participants did not start to identify as black after the experiment and so radically change their sense of self. Yet, having experienced the world in which this was the structured ontology, their implicit racial bias, and so their subjectivity, was changed by means of the intervention. As Gualeni and Vella (2020) write: "in virtual worlds, human beings can reflect on their values and beliefs, take on new subjectivities, explore previously unexperienced ways of being, and take reflective stances toward their existence and their subjectivity in the actual world." (p. xix).

The Relationship Between Environment and World

In addition to considering the relationship between Human and User in the framework, it can be fruitful to clarify the relationship between the Environment and the World. In the phenomenological tradition, a given world is not understood as equivalent with reality. Rather, a world is understood as how reality is disclosed by human beings (Verbeek, 2005a). Worlds are—in their intentional relationship to human beings—intelligible, persistent, and "understood together" (Gualeni, 2015). The virtual environment with its "world characteristics" is seen as a part of the regular world in which it is accessible; however, engagement with it leaves the "real world" in the background in the alterity relation that is constituted. Ihde (2002) describes alterity as a "quasi-other or quasi-world with which the human actor relates" (p. 81). The virtual environment can be quite "other": it does not need to behave according to traditional ontologies and can instead, as we have seen, inverse them. In short, virtual environments are "fictive world[s] that [are] constructed, not copied" (Ihde, 2002, p. 81) and they come with their own "integrated ontology" (Metzinger, 2018, p. 4). The point is, however, that although the real and virtual worlds have distinct self-contained ontologies of their own, they are nevertheless highly interrelated. Again, we return to the concept of mediation. Postphenomenology stresses the role that technologies have in mediating humans' intentional relation toward their world, and in the case of Immersive VR, it is the experience of a virtually structured ontology that might reframe how humans disclose their worlds, and vice versa. Thus, postphenomenologically, we understand the ontologies of VR and RL as interrelated, so that experiencing a differently structured ontology in VR might affect the ontology of one's real world, or as Gualeni (2015) formulates it; "people's capability for structuring thought and rationalizing experience in relation to the actual world." (p. 19).

The Scope of the Framework

This paper has presented a theoretical framework for understanding the user-environment relations that Immersive VR gives rise to. One may ask, however, whether the framework extends to other VR technologies such as desktop VR (e.g., computer games), Mixed Reality (MR) technologies, and Augmented Reality (AR) technologies. The identified user-environment relations we presented rests on the particular human-technology relation that VR constitutes: the possibility of embodying, as well as relating to as alterity, parts of the same technological mediator. Other immersive technologies, such as AR and MR, do not constitute the same human-technology relation as VR. They are mainly distinguished in that they are not so immersive, and therefore engagement with the world persists actively instead of existing as a background relation. MR, for instance, seem to constitute an immersion relation in the sense that the virtual is merged with the world, and so it is distinguished from Immersive VR in which there is not this "merging" of the physical and the virtual. AR technologies are also distinguished in the human-technology relation they constitute and are well described by Verbeek's augmentation relation. In the augmentation relation of a device such as the Google Glass, we embody the glasses, and we are in a hermeneutical relation to the technology, while our involvement with the world persists (Verbeek, 2015a).

The less immersive Desktop VR medium actually constitutes a similar relation to Immersive VR; the human interacts through an avatar toward the alterity of the virtual environment, where the world is in the background. Nevertheless, the experience is very different as Desktop VR is less immersive, and you can clearly see the bounds of the medium. The content is *framed*, and "[e]verything is in front of the participant" (Ihde, 2002, p. 10). This framing restricts the medium's capability to encapsulate the user, and so the Desktop VR cannot achieve the same kind of mimesis with reality that Immersive VR can, where user interfaces can be natural and transparent, and the mediated information appears as if non-mediated. The degree of isomorphism between reality and virtuality that a simulating medium can achieve is important because it dictates how objects with their horizons and affordances are available to the user. For instance, Immersive VR can enable user interfaces to utilize natural bodily engagement with the virtual world (e.g., physically jumping vs. pressing space, or rotating head vs. moving mouse). This is not to say that desktop VR interfaces cannot also be embodied, or that all Immersive VR applications utilize natural interaction exclusively. In terms of general medium characteristics, however, desktop VR is not as inherently intuitive as immersive VR and may require more time to embody properly, just as we need to learn to drive a car before it truly becomes an extension of our bodies and we can pay attention to the road rather than how to maneuver the car.

To conclude, AR and MR constitute different human-technology relations than Immersive VR, and so our framework of user-environment relations is not directly relevant for understanding user experience in environments using these technologies. The encapsulating capabilities of the Immersive VR medium distinguishes it from other computer simulation technologies like Desktop VR, which do not leave the world in the background to the same extent as Immersive VR technologies. The capability of Immersive VR to provide reality-based interaction also contribute to the differences in how we experience worlds mediated through Immersive VR as opposed to Desktop VR.

FUTURE WORK

The perspective of postphenomenology sees the technological research product as a mediator that gives rise to a particular user-environment relation. It purports the view that the design of technological objects should also be understood as the design of human subjects. As Willis (2006) posits in her idea of ontological design, having this understanding—that what we design also designs us—"inevitably means undertaking any kind of designing activity with a very different kind of disposition." (p. 82). Developing *systematic approaches* of incorporating this understanding in evaluation and anticipation, however, is outside the scope of this paper. Future work addressing the applicability of post-phenomenological theory to concrete, practical cases would therefore complement our research. Here, we wish to highlight two avenues for research as particularly promising. Firstly, the development of systematic approaches to the empirical study of user experience in VR, and secondly, the development of guidelines for anticipating mediations as part of design processes and ethical assessment. For empirically studying user experience in VR, we see contextual inquiries where users are interviewed/queried in the virtual environment (Schwind et al., 2017, 2019; Alexandrovsky et al., 2021) as promising venture points for understanding user-environment relations as such. In terms of anticipation, the post-phenomenological approach to "variational analysis" is highly relevant, which could be described as "brainstorming stabilities of a multi-stable technology" (ibid, p. 27). Rosenberger and Verbeek (2015) discuss how this approach is inspired by Husserl's eidetic reduction, but radically altered to find variations within particular contexts instead of the aim being to locate general "essences." Work looking into how postphenomenology's variational analyses can be performed more concretely for VR is here desirable.

Beyond the advancement of theoretical insights into methodology, however, what is most desired in future work is empirical insight into actual user-environment relations. In our analysis, we were not able to perform an analysis of the research participants' mediated experience, as in most of the cases, the participants' experiences were not outlined in-depth enough for it to be possible. Although assuming the participants' experienced the mediations as they were intended may be somewhat justified as the interventions were successful, we wish to stress that reaching experience through induction is not relating to actual, phenomenological accounts. In fact, the role of postphenomenology as we see it is precisely to move away from the researchers' assumptions of what experience is being mediated toward the actual mediated experience.

CONCLUSION

Immersive VR is a remarkably flexible medium for interventions as it allows the construction of virtual worlds with ontologies radically different from the real world. Moving toward an understanding of the experiences underlying these effective interventions, we have proposed a theoretical framework that sees the user experience in Immersive VR as mediated in relations constituted between user and environment. The perspective that we advocate is distinguished from traditional approaches to understanding user experience in that it does not presuppose the human subject and the technology as poles between which interaction occurs. Rather, it sees the human subject and the experienced technology as a *result* of this interaction and the user experience as mediated in relations constituted between user and environment. We purport this perspective is a more relevant way of understanding the user experience underlying VR's capability to "change the self," as it specifically attends to how the human subject is mediated in the user-environment relation that is constituted. The applicability of the framework has been demonstrated through an analysis of a variety of VR interventions that constitute particular user-environment relations that vary greatly in terms of their ontological structuring. Finally, we have discussed the interrelations of various aspects of our framework, addressed the framework's scope, and provided directions for future work in advancing the theoretical framework into the methodological.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

The article was written by JV. The contributions by BW include supervision, comments, presentation, and discussion of conceptual ideas. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by a Ph.D. stipend at the University of Bergen.

REFERENCES

Achterhuis, H. (2001). *American Philosophy of Technology: The Empirical Turn*. Bloomington, IN: Indiana University Press.

Ahn, S. J., Le, A. M. T., and Bailenson, J. (2013). The effect of embodied experiences on self-other merging, attitude, and helping behavior. *Media Psychol*. 16, 7–38. doi: 10.1080/15213269.2012.755877

Ahn, S. J. G., Bostick, J., Ogle, E., Nowak, K. L., McGillicuddy, K. T., and Bailenson, J. N. (2016). Experiencing nature: embodying animals in immersive virtual environments increases inclusion of nature in self and involvement with nature. *J. Comp. Mediat. Commun*. 21, 399–419. doi: 10.1111/jcc4.12173

Alaraj, A., Lemole, M., Finkle, J., Yudkowsky, R., Wallace, A., Luciano, C., et al. (2011). Virtual reality training in neurosurgery: Review of current status and future applications. *Surg. Neurol. Int*. 2:52. doi: 10.4103/2152-7806.80117

Alexandrovsky, D., Putze, S., Schwind, V., Mekler, E. D. S., Smeddinck, J. D., Kahl, D., et al. (2021). "Evaluating user experiences in mixed reality," in *CHI Conference on Human Factors in Computing Systems Extended Abstracts* (New York, NY), 5.

Amores, J., Fuste, A., and Richer, R. (2019). "Deep reality: towards increasing relaxation in VR by subtly changing light, sound and movement based on HR, EDA, and EEG," in *Extended Abstracts of the* 2019 *CHI Conference on Human Factors in Computing Systems* (New York, NY), 1–2.

Ates, H. C., Fiannaca, A., and Folmer, E. (2015). "Immersive simulation of visual impairments using a wearable see-through display," in *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction* (New York, NY), 225–228.

Aydin, C., González Woge, M., and Verbeek, P.-P. (2019). Technological environmentality: conceptualizing technology as a mediating milieu. *Philos. Technol*. 32, 321–338. doi: 10.1007/s13347-018-0309-3

Banakou, D., Beacco, A., Neyret, S., Blasco-Oliver, M., Seinfeld, S., and Slater, M. (2020). Virtual body ownership and its consequences for implicit racial bias are dependent on social context. *R. Soc. Open Sci*. 7:201848. doi: 10.1098/rsos.201848

Banakou, D., Hanumanthu, P. D., and Slater, M. (2016). Virtual embodiment of white people in a black virtual body leads to a sustained reduction in their implicit racial bias. *Front. Hum. Neurosci*. 10:601. doi: 10.3389/fnhum.2016.00601

Banakou, D., Kishore, S., and Slater, M. (2018). Virtually being einstein results in an improvement in cognitive task performance and a decrease in age bias. *Front. Psychol*. 9:917. doi: 10.3389/fpsyg.2018.00917

Bourdin, P., Barberia, I., Oliva, R., and Slater, M. (2017). A virtual out-of-body experience reduces fear of death. *PLoS ONE* 12:e0169343. doi: 10.1371/journal.pone.0169343

Çaliskan, O. (2011). Virtual field trips in education of earth and environmental sciences. *Proc. Soc. Behav. Sci*. 15, 3239–3243. doi: 10.1016/j.sbspro.2011.04.278

De Mul, J. (2010). *Cyberspace Odyssey: Towards a Virtual Ontology and Anthropology*. Newcastle upon Tyne: Cambridge Scholars Publishing.

Donohoe, J. (2017). "Introduction," in *Place and Phenomenology,* ed J. Donohoe (Washington, DC: RLI), 1–327.

Falconer, C. J., Slater, M., Rovira, A., King, J. A., Gilbert, P., Antley, A., et al. (2014). Embodying compassion: a virtual reality paradigm for overcoming excessive self-criticism. *PLoS ONE* 9:e111933. doi: 10.1371/journal.pone.0111933

Flobak, E., Wake, J. D., Vindenes, J., Kahlon, S., Nordgreen, T., and Guribye, F. (2019). "Participatory design of VR scenarios for exposure therapy," in *Proceedings of the* 2019 *CHI Conference on Human Factors in Computing Systems* (New York, NY), 1–12.

Fox, J., and Bailenson, J. N. (2009). Virtual self-modeling: the effects of vicarious reinforcement and identification on exercise behaviors. *Media Psychol*. 12, 1–25. doi: 10.1080/15213260802669474

Fox, J., Bailenson, J. N., and Tricase, L. (2013). The embodiment of sexualized virtual selves: the Proteus effect and experiences of self-objectification via avatars. *Comput. Human Behav*. 29, 930–938. doi: 10.1016/j.chb.2012.12.027

Gonzalez-Franco, M., Slater, M., Birney, M. E., Swapp, D., Haslam, S. A., and Reicher, S. D. (2018). Participant concerns for the Learner in a Virtual Reality replication of the Milgram obedience study. *PLoS ONE* 13:e0209704. doi: 10.1371/journal.pone.0209704

Groom, V., Bailenson, J. N., and Nass, C. (2009). The influence of racial embodiment on racial bias in immersive virtual environments. *Social Influence* 4, 231–248. doi: 10.1080/15534510802643750

Gualeni, S. (2015). *Virtual Worlds as Philosophical Tools: How to Philosophize With a Digital Hammer*. London: Palgrave Macmillan.

Gualeni, S., and Vella, D. (2020). *Virtual Existentialism: Meaning and Subjectivity in Virtual Worlds*. London: Springer Nature Switzerland, Palgrave Macmillan.

Hamilton-Giachritsis, C., Banakou, D., Garcia Quiroga, M., Giachritsis, C., and Slater, M. (2018). Reducing risk and improving maternal perspective-taking and empathy using virtual embodiment. *Sci. Rep*. 8:2975. doi: 10.1038/s41598-018-21036-2

Hasler, B. S., Spanlang, B., and Slater, M. (2017). Virtual race transformation reverses racial in-group bias. *PLoS ONE* 12:e0174965. doi: 10.1371/journal.pone.0174965

Hauser, S., Oogjes, D., Wakkary, R., and Verbeek, P.-P. (2018). "An annotated portfolio on doing postphenomenology through research products," in *Proceedings of the* 2018 *Designing Interactive Systems Conference* (New York, NY), 459–471.

Hershfield, H. E., Goldstein, D. G., Sharpe, W. F., Fox, J., Yeykelis, L., Carstensen, L. L., et al. (2011). Increasing saving behavior through age-progressed renderings of the future self. *J. Mark. Res*. 48, S23–S37. doi: 10.1509/jmkr.48.SPL.S23

Ihde, D. (1990). *Technology and the Lifeworld: From Garden to Earth*. Bloomington, IN: Indiana University Press.

Ihde, D. (2002). *Bodies in Technology*. Minneapolis, MN: University of Minnesota Press.

Ihde, D. (2003). Postphenomenology - Again? Available online at: https://sts.au.dk/fileadmin/sts/publications/working_papers/Ihde_-_Postphenomenology_Again.pdf

Ihde, D. (2012). *Experimental Phenomenology,* 2nd Edn. Albany, NY: State University of New York Press.

Introna, L. (2017). "Phenomenological approaches to ethics and information technology," in *Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University. Available online at: https://plato.stanford.edu/archives/fall2017/entries/ethics-it-phenomenology

Jones, P. R., Somoskeöy, T., Chow-Wing-Bom, H., and Crabb, D. P. (2020). Seeing other perspectives: evaluating the use of virtual and augmented reality to simulate visual impairments (OpenVisSim). *Npj Digital Med*. 3, 1–9. doi: 10.1038/s41746-020-0242-6

Kaposy, C. (2017). "Postphenomenology of the robot medical student," in *Postphenomenological Investigations: Essays on Human-Technology Relations*, eds R. Rosenberger and P.-P. Verbeek (Lanham, MD: Lexington Books), 191–201.

Kay, A., and Goldberg, A. (1977). Personal dynamic media. *Computer* 10, 31–41. doi: 10.1109/C-M.1977.217672

Lanier, J. (2017). *Dawn of the New Everything: Encounters With Reality and Virtual Reality,* 1st Edn. New York, NY: Henry Holt and Co., Inc.

Leitner, J. B., Ayduk, O., Mendoza-Denton, R., Magerman, A., Amey, R., Kross, E., et al. (2017). Self-distancing improves interpersonal perceptions and behavior by decreasing medial prefrontal cortex activity during the provision of criticism. *Soc. Cogn. Affect. Neurosci*. 12, 534–543. doi: 10.1093/scan/nsw168

Lindner, P., Miloff, A., Fagernäs, S., Andersen, J., Sigeman, M., Andersson, G., et al. (2019). Therapist-led and self-led one-session virtual reality exposure therapy for public speaking anxiety with consumer hardware and software: A randomized controlled trial. *J. Anxiety Disord*. 61, 45–54. doi: 10.1016/j.janxdis.2018.07.003

Loomis, J. M. (2016). Presence in virtual reality and everyday life: immersion within a world of representation. *Presence Teleoperat. Virtual Environ*. 25, 169–174. doi: 10.1162/PRES_a_00255

Madary, M., and Metzinger, T. K. (2016). Real virtuality: a code of ethical conduct. Recommendations for good scientific practice and the consumers of VR-technology. *Front. Robot. AI* 3, 1–23. doi: 10.3389/frobt.2016.00003

Maxhall, M., Backman, A., Holmlund, K., Hedman, L., Sondell, B., and Bucht, G. (2004). "Participants responses to a stroke training simulator," in *5th International Conference of Disability* (Oxford), 225–229.

McCarthy, J., and Wright, P. (2004). *Technology as Experience*. Cambridge, MA: The MIT Press.

Merleau-Ponty, M. (2002). *Phenomenology of Perception*. London; New York, NY: Routledge.

Metzinger, T. K. (2018). Why is virtual reality interesting for philosophers? *Front. Robot. AI* 5, 1–19. doi: 10.3389/frobt.2018.00101

Milgram, S. (1964). Group pressure and action against a person. *J. Abnorm. Soc. Psychol*. 69, 137–143. doi: 10.1037/h0047759

Moran, S., Jäger, N., Schnädelbach, H., and Glover, K. (2016). ExoPranayama: a biofeedback-driven actuated environment for supporting yoga breathing practices. *Pers. Ubiquitous Comp*. 20, 261–275. doi: 10.1007/s00779-016-0910-3

Neyret, S., Navarro, X., Beacco, A., Oliva, R., Bourdin, P., Valenzuela, J., et al. (2020). An embodied perspective as a victim of sexual harassment in virtual reality reduces action conformity in a later milgram obedience scenario. *Sci. Rep*. 10:6207. doi: 10.1038/s41598-020-62932-w

Nyre, L., and Vindenes, J. (2020). "Immersive journalism as witnessing," in *Immersive Journalism as Storytelling: Ethics, Production, and Design,* 1st ed, eds T. Uskali, A. Gynnild, S. Jones, and E. Sirkkunen (London: Routledge), 25–36.

Osimo, S. A., Pizarro, R., Spanlang, B., and Slater, M. (2015). Conversations between self and self as Sigmund Freud—a virtual body ownership paradigm for self counselling. *Sci. Rep*. 5:13899. doi: 10.1038/srep13899

Prpa, M., Cochrane, K., and Riecke, B. E. (2016). "Hacking alternatives in 21st century: designing a bio-responsive virtual environment for stress reduction," in *Communications in Computer and Information Science*, eds S. Serino, A. Matic, D. Giakoumis, G. Lopez, and P. Cipresso (Cham: Springer), 34–39.

Roo, J. S., Gervais, R., Frey, J., and Hachet, M. (2017). "Inner garden: connecting inner states to a mixed reality sandbox for mindfulness," in *Proceedings of the* 2017 *CHI Conference on Human Factors in Computing Systems* (New York, NY), 1459–1470.

Rosenberg, R. S., Baughman, S. L., and Bailenson, J. N. (2013). Virtual superheroes: using superpowers in virtual reality to encourage prosocial behavior. *PLoS ONE* 8:e55003. doi: 10.1371/journal.pone.0055003

Rosenberger, R., and Verbeek, P.-P. (2015). "A field guide to postphenomenology," in *Postphenomenological Investigations: Essays on Human-Technology Relations*, eds R. Rosenberger and P.-P. Verbeek (Lanham, MD: Lexington Books), 9–41.

Rua, H., and Alvito, P. (2011). Living the past: 3D models, virtual reality and game engines as tools for supporting archaeology and the reconstruction of cultural heritage – the case-study of the Roman villa of Casal de Freiria. *J. Archaeol. Sci*. 38, 3296–3308. doi: 10.1016/j.jas.2011.07.015

Satava, R. M. (1993). Virtual reality surgical simulator - the first steps. *Surg. Endosc*. 7, 203–205. doi: 10.1007/BF00594110

Schwind, V., Knierim, P., Haas, N., and Henze, N. (2019). "Using presence questionnaires in virtual reality," in *Proceedings of the* 2019 *CHI Conference on Human Factors in Computing Systems* (New York, NY), 1–12.

Schwind, V., Knierim, P., Tasci, C., Franczak, P., Haas, N., and Henze, N. (2017). ""These are not my hands!": effect of gender on the perception of avatar hands in virtual reality," in *Proceedings of the* 2017 *CHI Conference on Human Factors in Computing Systems* (New York, NY), 1577–1582.

Seinfeld, S., Arroyo-Palacios, J., Iruretagoyena, G., Hortensius, R., Zapata, L. E., Borland, D., et al. (2018). Offenders become the victim in virtual reality: impact of changing perspective in domestic violence. *Sci. Rep*. 8:2692. doi: 10.1038/s41598-018-19987-7

Semertzidis, N. A., Sargeant, B., Dwyer, J., Mueller, F. F., and Zambetta, F. (2019). "Towards understanding the design of positive pre-sleep through a neurofeedback artistic experience," in *Proceedings of the* 2019 *CHI Conference on Human Factors in Computing Systems* (New York, NY), 1–14.

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philos. Transac. R. Soc. B Biol. Sci*. 364, 3549–3557. doi: 10.1098/rstb.2009.0138

Slater, M., Antley, A., Davison, A., Swapp, D., Guger, C., Barker, C., et al. (2006). A virtual reprise of the stanley milgram obedience experiments. *PLoS ONE* 1:e39. doi: 10.1371/journal.pone.0000039

Slater, M., Gonzalez-Liencres, C., Haggard, P., Vinkers, C., Gregory-Clarke, R., Jelley, S., et al. (2020). The ethics of realism in virtual and augmented reality. *Front. Virtual Real*. 1, 1–13. doi: 10.3389/frvir.2020.00001

Slater, M., and Sanchez-Vives, M. V. (2014). Transcending the self in immersive virtual reality. *Computer* 47, 24–30. doi: 10.1109/MC.2014.198

Slater, M., and Sanchez-Vives, M. V. (2016). enhancing our lives with immersive virtual reality. *Front. Robot. AI*. 3, 1–47. doi: 10.3389/frobt.2016.00074

Sri Kalyanaraman, S., Penn, D. L., Ivory, J. D., and Judge, A. (2010). The virtual Doppelganger. *J. Nerv. Mental Dis*. 198, 437–443. doi: 10.1097/NMD.0b013e3181e07d66

Stepanova, E. R., Desnoyers-Stewart, J., Pasquier, P., and Riecke, B. E. (2020). "JeL: breathing together to connect with others and nature," in *Proceedings of the* 2020 *ACM Designing Interactive Systems Conference* (New York, NY), 641–654.

Suzuki, K., Roseboom, W., Schwartzman, D. J., and Seth, A. K. (2017). A deepdream virtual reality platform for studying altered perceptual phenomenology. *Sci. Rep*. 7:15982. doi: 10.1038/s41598-017-16316-2

Tajadura-Jiménez, A., Banakou, D., Bianchi-Berthouze, N., and Slater, M. (2017). Embodiment in a child-like talking virtual body influences object size perception, self-identification, and subsequent real speaking. *Sci. Rep*. 7:9637. doi: 10.1038/s41598-017-09497-3

van Loon, A., Bailenson, J., Zaki, J., Bostick, J., and Willer, R. (2018). Virtual reality perspective-taking increases cognitive empathy for specific others. *PLoS ONE* 13:e0202442. doi: 10.1371/journal.pone.0202442

Verbeek, P.-P. (2005a). *What Things Do (Philosophical Reflections on Technology, Agency, and Design)*, 1st Edn. University Park, MD: Penn State University Press.

Verbeek, P.-P. (2005b). "Beyond the human eye. Mediated vision and posthumanity," in *Proceedings of AIAS Conference "Mediated Vision"*, eds P. J. H. Kockelkoren and P. Kockellkoren. (Veenman Publishers en ARTez Press), 1–7. Available online at: https://research.utwente.nl/en/publications/beyond-the-human-eye-mediated-vision-and-posthumanity

Verbeek, P.-P. (2011). *Moralizing Technology: Understanding and designing the Morality of Things*. Chicago, IL: University of Chicago Press.

Verbeek, P.-P. (2015a). "Designing the public sphere: information technologies and the politics of mediation," in *The Onlife Manifesto*, ed L. Floridi (Cham: Springer International Publishing), 217–227.

Verbeek, P.-P. (2015b). Cover story: beyond interaction: a short introduction to mediation theory. *Interactions* 22, 26–31. doi: 10.1145/2751314

Verbeek, P. P. (2008). "Cyborg intentionality: Rethinking the phenomenology of human-technology relations," in *Phenomenology and the Cognitive Sciences*, ed S. Evan (Luxembourg: Springer Science & Business Media B.V), 387–395.

Voordijk, H., and Vahdatikhaki, F. (2020). Virtual Reality learning environments and technological mediation in construction practice. *Eur. J. Eng. Educ*. 1–15. doi: 10.1080/03043797.2020.1795085

Willis, A.-M. (2006). Ontological designing. *Design Philos. Pap*. 4, 69–92. doi: 10.2752/144871306X13966268131514

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Virtual Reality

SYSTEMATIC REVIEW
published: 14 July 2021
doi: 10.3389/frvir.2021.694567

A Systematic Review of Physiological Measurements, Factors, Methods, and Applications in Virtual Reality

***Andreas Halbig** * and Marc Erich Latoschik*

Human-Computer Interaction (HCI) Group, Informatik, University of Würzburg, Würzburg, Germany

Measurements of physiological parameters provide an objective, often non-intrusive, and (at least semi-)automatic evaluation and utilization of user behavior. In addition, specific hardware devices of Virtual Reality (VR) often ship with built-in sensors, i.e. eye-tracking and movements sensors. Hence, the combination of physiological measurements and VR applications seems promising. Several approaches have investigated the applicability and benefits of this combination for various fields of applications. However, the range of possible application fields, coupled with potentially useful and beneficial physiological parameters, types of sensor, target variables and factors, and analysis approaches and techniques is manifold. This article provides a systematic overview and an extensive state-of-the-art review of the usage of physiological measurements in VR. We identified 1,119 works that make use of physiological measurements in VR. Within these, we identified 32 approaches that focus on the classification of characteristics of experience, common in VR applications. The first part of this review categorizes the 1,119 works by field of application, i.e. therapy, training, entertainment, and communication and interaction, as well as by the specific target factors and variables measured by the physiological parameters. An additional category summarizes general VR approaches applicable to all specific fields of application since they target typical VR qualities. In the second part of this review, we analyze the target factors and variables regarding the respective methods used for an automatic analysis and, potentially, classification. For example, we highlight which measurement setups have been proven to be sensitive enough to distinguish different levels of arousal, valence, anxiety, stress, or cognitive workload in the virtual realm. This work may prove useful for all researchers wanting to use physiological data in VR and who want to have a good overview of prior approaches taken, their benefits and potential drawbacks.

Keywords: virtual reality, use cases, sesnsors, tools, biosignals, psychophyisology, HMD (Head-Mounted Display), systematic review

OPEN ACCESS

Edited by:
Missie Smith, Independent Researcher, United States

Reviewed by:
Frank Guan, Singapore Institute of Technology, Singapore
Mariano Alcañiz, Universitat Politècnica de València, Spain

****Correspondence:***
Andreas Halbig
andreas.halbig@uni-wuerzburg.de

Specialty section:
This article was submitted to Virtual Reality and Human Behaviour, a section of the journal Frontiers in Virtual Reality

Received: *13 April 2021*
Accepted: *17 June 2021*
Published: *14 July 2021*

Citation:
Halbig A and Latoschik ME (2021) A Systematic Review of Physiological Measurements, Factors, Methods, and Applications in Virtual Reality. Front. Virtual Real. 2:694567. doi: 10.3389/frvir.2021.694567

1 INTRODUCTION

Virtual Reality (VR) provides the potential to expose people to a large variety of situations. One advantage it has over the exposure to real situations is that the creator of the virtual environment can easily and reliably control the stimuli that are presented to an immersed person (Vince, 2004). Usually, the presented stimuli are not arbitrary but intentionally chosen to evoke a certain experience in the user, e.g. anxiety, relaxation, stress, or presence. Researchers require tools that help them to determine whether the virtual environment fulfills its purpose and how users respond to certain stimuli. Evaluation methods are an essential part of the development and research of VR.

Evaluation techniques can be divided into implicit and explicit methods (Moon and Lee, 2016; Marín-Morales et al., 2020). Explicit methods require the user to *explicitly* and actively express the own experience. Hence, they can also be called subjective methods. Examples include interviews, thinking-aloud and questionnaires. In the evaluation of VR, questionnaires are the most prominent explicit method. They are very versatile and designed for the quantification of various characteristics of experience. Some assess VR specific phenomena, e.g. presence (Slater et al., 1994; Witmer and Singer, 1998), simulator sickness (Kennedy et al., 1993), or the illusion of virtual body-ownership (Roth and Latoschik, 2020). Other questionnaires capture more generic characteristics of experience, but are still useful in many VR scenarios, e.g. workload (Hart and Staveland, 1988) or affective reactions (Watson et al., 1988; Bradley and Lang, 1994).

Traditionally, questionnaires and other explicit methods also bring with them some disadvantages. There is a variety of self-report biases that can manipulate the way people respond to questions. A common example is the social desirability bias. It refers to the idea that subjects tend to choose a response that they expect to meet social expectations instead of one that reflects their true experience (Corbetta, 2003; Grimm, 2010). Other common examples for response biases are the midpoint bias where people tend to choose neutral answers (Morii et al., 2017) or extreme responding where people tend to choose the extreme choices on a rating scale (Robins et al., 2009). In general, there is a variety of characteristics and circumstances that can negatively influence the human capacity to evaluate oneself. For a detailed description of erroneous self-assessment of humans, refer to Dunning et al. (2004). Another point that can limit the validity of questionnaires is that one never knows if the questions were understood by participants (Rowley, 2014). The complexity of the information and the language skills of the respondents can influence how questions are interpreted and thus how answers turn out (Redline et al., 2003; Richard and Toffoli, 2009). Another problem is that explicit methods often separate the evaluation from the underlying stimulus. Thus, they rely on a correct recapitulation of experience. People might not be able to remember how exactly they were feeling when they were interacting with a software (Cairns and Cox, 2008). This is especially relevant for VR, as leaving the virtual environment can lead to a change in the evaluation of the experience (Schwind et al., 2019). In addition, some mental processes are not even accessible to consciousness and are therefore not recorded by explicit methods (Barsade et al., 2009).

Implicit evaluation methods avoid a lot of those drawbacks. In contrast to the explicit measures, they do not require the user to actively participate in the evaluation. Rather, they analyze the user behavior based on the response to a certain stimulus or event. This can be done either by direct observation or by analysis of physiological data. These implicit methods can also be referred to as objective methods as they do not rely on the ability of subjects to assess their own condition. Implicit evaluation that is based on physiological data has the advantage that it can assess both, automatic and deliberate processes. With automatic processes, we refer to organic activations that are unconsciously controlled by the autonomous nervous system, e.g. bronchial dilation or the activation of the sweat secretion (Jänig, 2008; Laight, 2013). These activations cannot be observed from the outside. With measures like electrodermal activity, electrocardiography, or electroencephalography, however, we can assess them. This allows quantification of how current stimuli are processed by the nervous system (Jänig, 2008; Laight, 2013). Deliberate processes, on the other hand, do not depend on unconscious activations of the autonomous nervous system. Nevertheless, physiological measures can help to understand these processes. Electromyography, for example, measures the strength of contraction of skeletal muscles (De Luca, 2006). Thus, this signal can also depend on arbitrary control by the human being.

Physiological measurements offer decisive advantages. They can be taken during exposure, they do not depend on memory, they can capture sub-conscious states, data can be collected fairly unobtrusively, and they yield quantitative data that can be leveraged for machine-learning approaches. A depiction of the discussed structure of evaluation methods can be found in **Figure 1**. An overview of the physiological measures that are considered in this work can be found in **Table 1**. It also contains abbreviations for the measurements that are used from now on.

The availability of easy-to-use wearable sensors is spurring the use of physiological data. EEG headsets such as the EPOC+[1] or the Muse 2,[2] wrist and chest worn trackers from POLAR,[3] fitbit,[4] and Apple[5]; as well as the EMPATICA E4[6] all make it easier to collect physiological data. In addition, VR headsets already come with built-in sensors that can be used for behavior analysis. Data from gyroscopes and accelerometers, included in VR-headsets and controllers, provide direct information about movement patterns. Moreover, eye-tracking devices from tobii[7] or Pupil Labs[8] can be used to easily extend VR-headsets so they deliver even more data, e.g. pupil dilation or blinking rate.

Due to the aforementioned advantages in combination with the availability of easy-to-use sensors and low-cost head-mounted displays (HMD) (Castelvecchi, 2016), the number of research approaches that combine physiological data with VR has increased considerably in recent years. In their meta-review about emotion recognition in VR with physiological data, Marín-Morales et al. (2020) even report an exponential growth of this field. Researchers who want to use physiological measures for their own VR application, however, are faced with a very rapidly growing field that offers a wide range of possibilities. As previously implied, there are a variety of signals that can be collected with a variety of sensors. While it is clear that a presence

[1]https://www.emotiv.com/epoc/
[2]https://choosemuse.com/de/muse-2/
[3]https://www.polar.com/
[4]https://www.fitbit.com/global/de/home
[5]https://www.apple.com/watch/
[6]https://www.empatica.com/research/e4/
[7]https://www.tobii.com
[8]https://pupil-labs.com

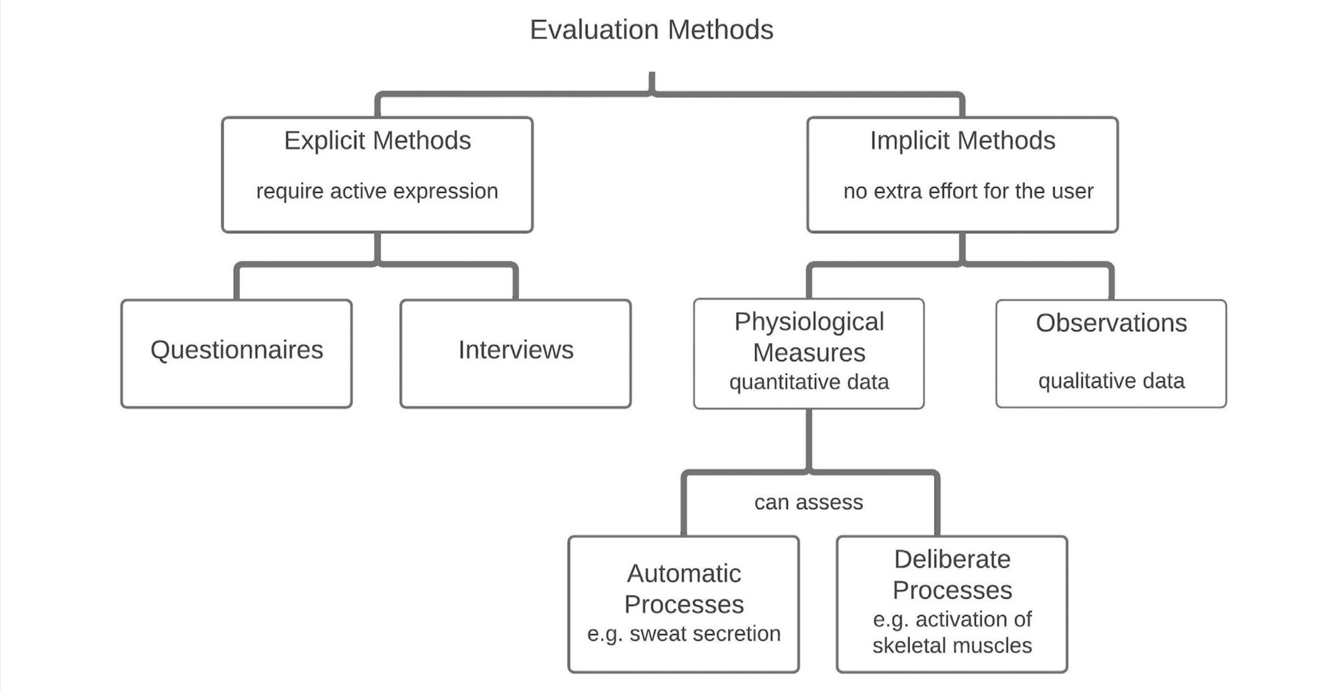

FIGURE 1 | Categorization of evaluation methods. The overview should not be considered comprehensive, but mereley as an orientation.

questionnaire is used to assess presence, such a 1-to-1 linkage of measure and experience is not possible for physiological measures. Their usage in VR applications is therefore anything but trivial. A structured reappraisal of the field is necessary.

This systematic review consists of two parts that address the following issues:

In the first part of this article, we examine the different use cases of physiological measurements in VR. We collect a broad selection of works that use physiological measures to assess the state of the user in the virtual realm. Then we categorize the works into specific fields of application and explain the functionality of the physiological measurements within those fields. As a synthesis of this part, we describe a list of the main purposes of using physiological data in VR. This serves as a broad, state of the art overview of how physiological measures can be used in the field of VR.

However, knowing what this data can be used for is only half the battle. We still need to know how to work with this data, in order to gain knowledge about a user's experience. Hence, in the second part of this paper, we will discuss concrete ways to collect and interpret physiological data in VR. Works that tell us a lot about how to get data and what can be deduced from it are classification approaches. To be precise, this includes approaches, that classify different levels of certain characteristics of experience. The works from this domain usually adopt the characteristics of experiences as their independent variable. Subjects in those studies were exposed to stimuli known to elicit a certain experience, such as anxiety. These studies then examined the extent to which the change in experience was reflected in physiological measurements. Thus, the works focus on the physiological measures themselves and their ability to quantify a particular experience. This review of classifiers, therefore, provides a clear overview of signals, sensors, tools and algorithms, that have been sensitive enough to distinguish different levels of the targeted experience in a VR setup. They show concrete procedures on how to extract the information that is hidden in the physiological data.

2 METHODS

On the basis of the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement (Liberati et al., 2009), we searched and assessed literature to find papers that make use of physiological data in VR. We always searched for one specific signal in combination with VR, so the search terms consisted of two parts that are connected with an AND. Thus, the individual search terms can be summarized in one big query that can be described like this: (*"Virtual Reality" OR "Virtual Environment" OR "VR" OR "HMD") AND (Pupillometry OR "Pupil* Size" OR "Pupil* Diameter" OR "Pupil* Dilation" OR "Pupil" OR "Eye Tracking" OR "Eye-Tracking" OR "Eye-Tracker*" OR "Gaze Estimation" OR "Gaze Tracking" OR "Gaze-Tracking" OR "Eye Movement" OR "EDA" OR "Electrodermal Activity" OR "Skin Conductance" OR "Galvanic Skin Response" OR "GSR" OR "Skin Potential Response" OR "SPR" OR "Skin Conductance Response" OR "SCR" OR "EMG" OR "Electromyography" OR "Muscle Activity" OR "Respiration" OR "Breathing" OR "Heart Rate" OR "Pulse" OR "Skin Temperature" OR "Thermal Imaging" OR "Surface Temperature" OR "Blood Pressure" OR "Blood Volume Pressure" OR "EEG" OR "Electroencephalography"*). The terms had to be included in the title, abstract, or

TABLE 1 | Physiological measures that are commonly used in VR applications.

Measures	Components and features	Description	General use
Electrocardiogram (ECG) Photoplethysmogram (PPG) Blood Pressure (BP)	Heart Rate (HR), Heart Rate Variability (HRV), Blood Volume Pulse (BVP), Low-Frequency power, High-Frequency power	Measurement of the cardiovascular activity through electrical (ECG), optical (PPG), or pressure (BP) sensors.	Used primarily in medicine to monitor heart health. It can indicate defects in the heart function, e.g. in case of heart attacks or cardiac arrhythmia, and is thus often used as a diagnostic tool (Vanderlei et al., 2009; Alian and Shelley, 2014).
Electrodermal Activity (EDA)/Galvanic Skin Response (GSR)	Skin Conductance Level (SCL), Skin Potential Response (SPR), Mean, Standard Deviation, Peaks	Measuring electrical properties of the skin, which are most commonly influenced by sweat secretion.	Commonly used to detect changes in sympathetic activity caused by emotional and cognitive processing. Therefore it is often applied to gauge stress responses and emotional reactions (Benedek and Kaernbach, 2010; Braithwaite et al., 2013).
Skin Temperature (SKT)	Mean, Minimum, Maximum,	Measuring the surface temperature on certain areas of the skin.	Usually used to detect illness. When realized through thermal cameras, it can serve as a non-contact assessment of autonomous nervous activity (Kataoka et al., 1998).
Respiration (RESP)	Respiration Rate, Tidal Volume, O_2 Consumption, Mean, Standard Deviation, Peaks	Measurement of the breathing activity.	An important vital sign that is very good at predicting serious clinical events and identifying patients at risk. It is also often used to measure physiological load (Masaoka and Homma, 1997; AL-Khalidi et al., 2011).
Electromyogram (EMG)	Mean, Standard Deviation, Minimum, Maximum	Measuring the strength of the contraction of skeletal muscles, based on electrical signals.	Primarily used for studying human movement, the diagnosis of neuromuscular diseases, and for the active control of artificial limbs (Pullman et al., 2000; De Luca, 2006).
Eye-Tracking (ET)	Gaze, Fixations, Saccades, Pupil Dilation, Blink rate	Measuring eye movement and pupil properties, usually with cameras pointing towards the eyes.	Usually used as an estimation on which object the gaze falls on. This knowledge is often used in marketing research, usability tests, and human-computer interaction. Changes in pupil diameter can indicate cognitive processing (Singh and Singh, 2012; Sirois and Brisson, 2014).
Electroencephalogram (EEG)	Frequency Bands (Alpha, Beta, Gamma, Delta, and Theta), Mean Bandpower, Event Related Potential, Mean Amplitude	Measuring the electrical activity of the brain on the scalp.	Used in medicine for the diagnosis of neurological diseases such as epilepsy and seizures. In human-computer interaction, it is commonly used for brain-computer interfaces (Lai et al., 2018; Lotte et al., 2018).

keywords of an article. Queried databases were ACM Digital Library, Web of Science, PubMed, APA PsycInfo, PsynDex, and IEEE Xplore. The date of the search was October 15, 2020.

We gathered the results and inserted them into a database together with some extra papers that were known to be relevant for the topic. We removed duplicates and then started with the screening process. Papers were excluded if they were from completely different domains (VR, for example, can not only stand for "Virtual Reality"), if new sensors or algorithms were only introduced (but not actually used), if they only dealt with augmented reality, or if they were just presenting the idea of using physiological data and VR (but not actually did it). Furthermore, we excluded poster presentations, abstracts, reviews, and works that were not written in English. That means, left after this screening process were all works that present a use case in which the sought-after physiological measures were used together with a VR application. We usually screened papers on basis of title and abstract. About 10% required inspection of the full text to determine if they met the criteria. If the full text of those works was not available they were also excluded. We used the papers that were left after this process for the first part of this work. During the screening process, we began to note certain repetitive fields of application and compiled a list of categories and sub-categories of field of application. We then tagged the papers with these categories according to their field of application. We also noted the purpose for which the physiological data was used, also with the help of tags.

As already explained in the introduction (**Section 1**), in the further course of the review we focused on classification approaches. During the aforementioned tagging process, we identified papers that deal with some kind of classification. Those papers were then examined for their eligibility to be included in the second part of the review. The criterion for the inclusion of a paper here was that it is a work that

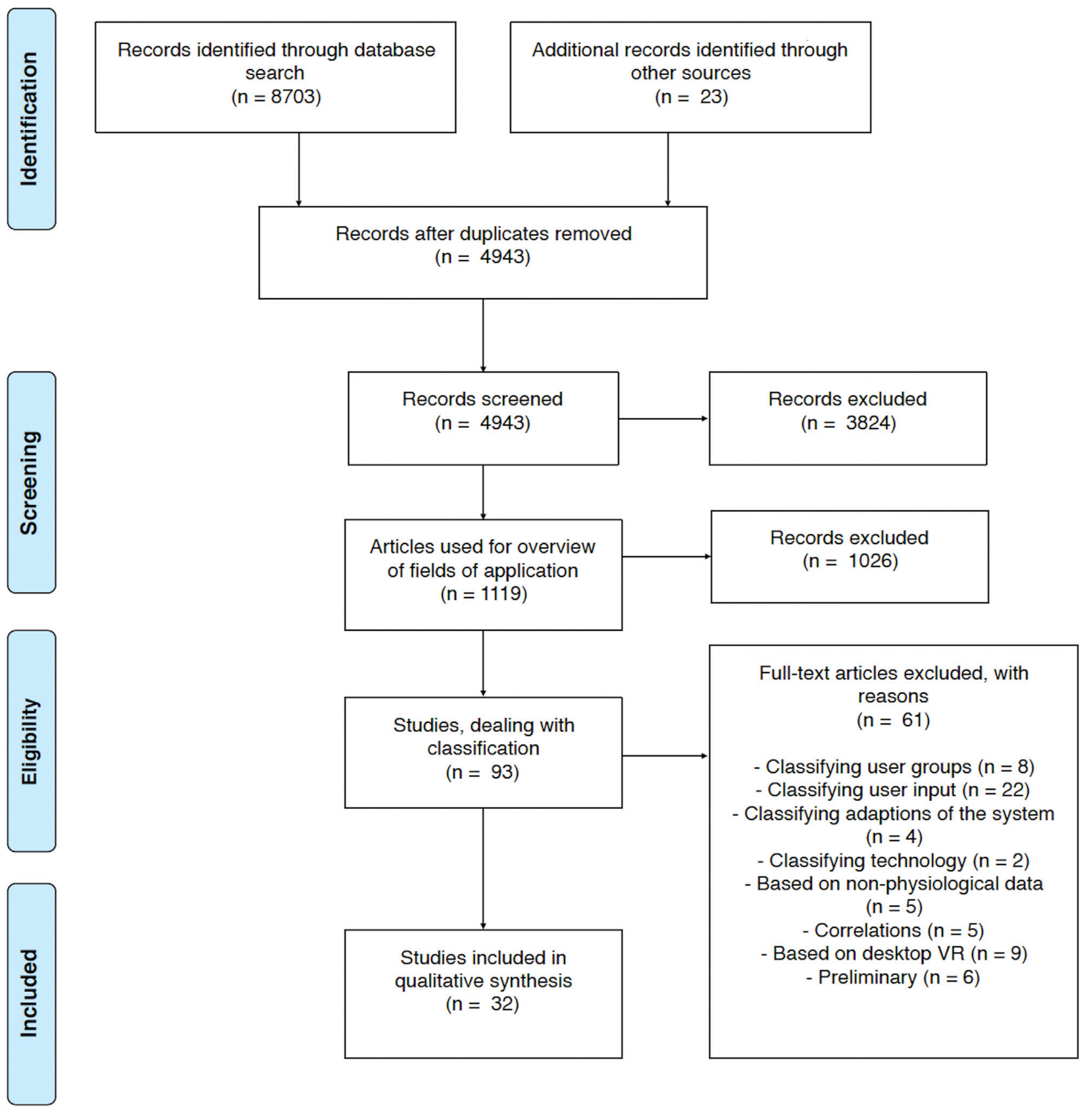

FIGURE 2 | Process and Results of the Literature Review. Diagram is adapted from Liberati et al. (2009).

presents a classification based on physiological data which was captured during exposure to immersive VR (CAVE-based or HMD-based). In order to be included, the work also had to distinguish different levels of current experience (e.g. high vs. low stress) and not different groups of people (e.g. children with and without ADHD). Excluded were classifiers that aim at the recognition of user input, an adaption of the system, or the recognition of the used technology. Also excluded were works that just look for correlations between signals and certain events, classification approaches that are based on desktop VR or non-physiological data.

3 RESULTS

An overview of the specific phases of the search for literature and the results can be seen in **Figure 2**. In total, the literature research yielded 4,943 different works. After the first screening process, 1,408 works were left over. They all show examples of how

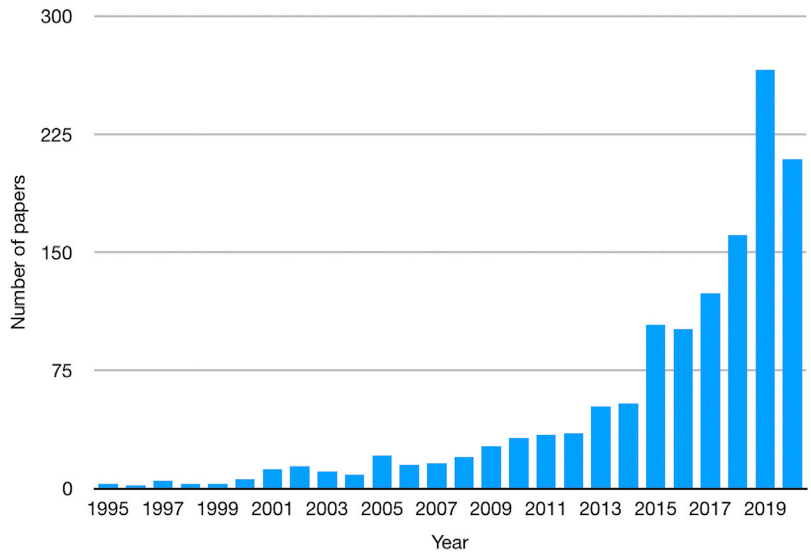

FIGURE 3 | Division of the 1,408 papers, that remained after screening, according to published year.

physiological data can be used in immersive VR to assess the state of the user. **Figure 3** shows the distribution of the papers over the years since 1995.

From this point forward, we chose to continue with works from 2013 and later. Thus, we shifted the focus to current trends. The numbers show that most of the papers were published during the last years (see **Figure 3**). The year 2013 was the first year for which we found more than 50 papers. This left 1,119 of the 1,408 papers that were published in 2013 or later. During the screening process, described in **Section 2**, we identified five major fields of application to which most works can be assigned to. These domains are therapy and rehabilitation, training and education, entertainment, functional VR properties and general VR properties. In the first part of the discussion section, we use this domain division to give a broad overview of the usage of physiological measures in VR (see **Section 4.1**).

After screening and checking for eligibility, 32 works that deal with the classification of experience in VR were left for further qualitative analysis. Each of the 32 works use physiological measures as dependent variables. As independent variables they manipulate the intensity of a target characteristic of experience. Thus, the works show the extent to which the physiological measures were able to reflect the manipulation of the independent variable. In our results, the most commonly assessed characteristic of experience was arousal, used in nine works, followed by valence and anxiety, both used in six works. Five works classify stress, and four, cognitive workload. The following characteristics of experience were measured in only one work: Visual fatigue, moments of insight, cybersickness, and understanding. An overview of the 32 works can be found in **Table 2**. This overview shows which characteristics of experience the works assessed, which measures and sensors they used for their approach, and which classification algorithms were chosen for the interpretation of the data. **Table 3** provides a separate overview of the sensors that were used in the 32 works. In **Table 4** we list different tools that were used in various classification approaches to record, synchronize, and process the physiological data. In the second part of the discussion, we deal with the listed characteristics of experience individually and summarize the corresponding approaches with a focus on signals and sensors (see **Section 4.2**).

4 DISCUSSION

As already indicated in **Section 3** the discussion of this work is divided into two parts.

4.1 Part 1: Fields of Application for Physiological Data in Virtual Reality

In the first part, we give a categorized overview of the copious use cases of physiological measures in VR. This overview is based on the 1,119 works and the fields of application that we identified during the screening process. This section is structured according to those fields. We highlight which works belong to which fields and how physiological measures are used. We summarize this overview by listing meta purposes for which physiological data are used in VR. This overview cannot cover all the works that are out there. What we describe are the types of work that have occurred most frequently. After all, this is an abstraction of the field. With only a few exceptions, all the examples we list here are HMD-based approaches.

4.1.1 Therapy and Rehabilitation

Therapy and rehabilitation applications are frequent fields of application for physiological measurements. Here, we talk about

TABLE 2 | Overview of classification approaches based on physiological data collected in full-immersive VR. If not stated otherwise, the values presented in the Results column usually refer to the accuracy that was achieved in a cross validation or on an extra test-set. These values serve only as a rough guide to the success of the method and are not comparable 1-to-1.

Study	Classification of	Measures and sensors	Subjects	Data acquisition	Algorithm	No. Classes	Results
Siravenha et al. (2019)	Cognitive Workload	EEG - 19 electrode headset (BrainMaster)	36	Mining task in an excavator simulator and driving simulator	MLPN	2	80.69% (test-set)
Xu et al. (2019)	Valence	EEG - 3 textile dry electrodes (custom)	19	4 different affective scenes	SVM	2	81.30% (test-set)
Orlosky et al. (2019)	Understanding (of words)	ET - Integrated camera (Pupil Labs)	16	Virtual word-recall task	SVM	2	75.60% (full cv)
Salkevicius et al. (2019)	Anxiety	PPG - Wristband (Empatica) EDA - Wristband (Empatica) SKT - Wristband (Empatica)	30	Speech in front of a real and a virtual audience	SVM	4	86.30% (10× 10-fold cv)
Bălan et al. (2020)	Anxiety	EEG - Cap with 16 dry electrodes (Brain Products) PPG - Finger photo diode (Shimmer) EDA - Finger electrodes (Shimmer)	8	Virtual and *in vivo* exposure to different heights	DNN kNN	2 4	89.50% (test-set) 52.75% (test-set)
Jeong et al. (2019)	Cybersickness	EEG - 14 channel headset (Emotiv)	24	Various 360° videos (violently moving, tranquil, or scary)	DNN	2	98.82 (3-fold-cv)
Cho et al. (2017)	Stress	PPG - 1 finger photo diode (Biopac) EDA - 2 finger electrodes (Biopac) SKT - 1 finger sensor (Biopac)	12	Videos (relaxing and dynamic) + arithmetic tasks	K-ELM	5	≈95% (loocv)
Tremmel et al. (2019)	Cognitive Workload	EEG - Cap with 8 wet electrodes (Ladybird)	15	N-back task with colored balls	LDA	2 3	81.1% (4-fold cv) 63.9% (4-fold cv)
Collins et al. (2019)	Cognitive Workload	PPG - Wristband (Empatica)	24	4D cube puzzles	RF	3	91.75% (10-fold cv)
Collins et al. (2019)	Moment of insight (Aha! moment)	EDA - Wristband (Empatica)	24	4D cube puzzles	RF	2	98.81% (10-fold cv)
Kakkos et al. (2019)	Cognitive Workload	EEG - Cap with 64 Ag/AgCl electrodes (antneuro)	33	Flight simulator task with three different difficulties	LDA	3	89.00% (100× 10-fold cv)
Ishaque et al. (2020)	Stress	ECG - 2 thorax electrodes EDA - 2 finger electrodes RESP- 2 abdomen electrodes	14	Virtual Roller coaster + Stroop task, relaxation game	GB	2	85.00% (5-fold cv)
Ham et al. (2017)	Stress	PPG - 1 finger photo diode (Biopac)	6	Static relaxing video (beach), dynamic scary video (patrolling guard)	LDA	3	≈79.00% (10-fold cv)
Tartarisco et al. (2015)	Stress	ECG - Chest band (PBS) RESP - Chest band (PBS) Motion - Chest band (PBS)	20	Stressfull work scenarios for a nurse	SOM	4	≈83.00% (loocv)
Ding Y. et al. (2020)	Arousal	EEG - 4-channel headband (interaxon)	18	Relaxation scene (nature), stone dodging game	CNN	2	86.03% (3-fold cv)
Hofmann et al. (2018)	Arousal	EEG - Cap with 30 Ag/AgCL electrodes (Brain Products)	45	Virtual roller coaster, Breaks	LSTM	2	75.70% (10-fold cv)
Mavridou et al. (2018b)	Arousal	PPG - two sensors integrated in HMD (emteq) ECG - Chest band (custom)	11	Affective short videos	SVM	2	0.69 (Area under ROC curve)
Shumailov and Gunes (2017)	Arousal, Valence	EMG - 8-channel armband (Thalmic Labs)	7	A variety of VR games	SVM	2 2	0.91 (F1)(Arousal) 0.85 (F1)(Valence)
Teo and Chia (2018)	Arousal/Excitement	EEG - 4-channel headband (interaxon)	24	Virtual roller coaster, Breaks	DNN	2	96.32% (10x10-fold cv)
Zheng et al. (2020)	Arousal, Valence	ET - Integrated camera (Pupil Labs)	10	Emotional 360° videos	SVM	4	57.05%
Marín-Morales et al. (2018)	Arousal, Valence	EEG - 9 electrode head strip (Biopac) ECG - 2 electrodes (placed on rib and collarbone)(Biopac)	60	Different architectural environments	SVM	2 2	75.00% (loocv) (Arousal) 71.21% (loocv) (Valence)
Bilgin et al. (2019)	Arousal	EEG - 4-channel headband (interaxon)	10	Calm environment (nature) and virtual roller coaster	SVM	2	66.88% (10-fold cv)
Suhaimi et al. (2020)	Arousal, Valence	EEG - 4 channel headband (interaxon)	31	Emotional 360° videos	RF	4	82.49% (10-fold cv)

(Continued on following page)

TABLE 2 | (*Continued*) Overview of classification approaches based on physiological data collected in full-immersive VR. If not stated otherwise, the values presented in the Results column usually refer to the accuracy that was achieved in a cross validation or on an extra test-set. These values serve only as a rough guide to the success of the method and are not comparable 1-to-1.

Study	Classification of	Measures and sensors	Subjects	Data acquisition	Algorithm	No. Classes	Results
Hu et al. (2018)	Anxiety	EEG - Cap with 30 scalp electrodes (Neuroscan) EOG - 2 scalp electrodes (Neuroscan)	60	Standing on the ground, standing on a plank	CNN	4	88.77% (10-fold cv)
Kaur et al. (2019)	Anxiety	EEG - Cap with 58 electrodes (Brain Products) EOG - 6 scalp electrodes (Brain Products)	10	Body leaning task on elevated ground	kNN	2	0.85 (F1) (5-fold cv)
Wang et al. (2018)	Anxiety	EEG - Cap with 30 Ag/AgCl electrodes (Neuroscan) EOG - 4 Ag/AgCl scalp electrodes (Neuroscan)	76	Standing on the ground, standing on a plank in high altitude	SVM	3	96.20% (5-fold cv)
Wang Y. et al. (2019)	Visual fatigue	ET - Integrated camera (7invensun)	105	Watching VR videos	SVM	2 3 4	90.79% (cv) 79.47% (cv) 74.25% (cv)
Robitaille and McGuffin (2019)	Stress	ECG - Chest band (Polar) Motion - Hand controllers and Motion Trackers	12	Hitting moving targets in a calm or uncanny environment	DT	2	81.10% (10× 2 cv)
Tremmel (2020)	Cognitive Workload	EEG - Cap with 8 wet electrodes (Ladybird) Motion - VR headset and controller	15	N-back task with colored balls	ANN	3	84.30% (5-fold cv)
Mavridou et al. (2018a)	Valence	EMG - 8 electrodes integrated in HMD (emteq)	34	Affective video content	SVM	3	82.50% (loocv)
Balan et al. (2019)	Anxiety	EEG - Cap with 16 dry electrodes (Brain Products) PPG - Finger photo diode (Shimmer) EDA - Finger electrodes (Shimmer)	4	Virtual and *in vivo* exposure to different heights	DNN	2 4	72.90% (test-set) 41.89% (test-set)

MLPN, Multilayer Percpetron Network; SVM, Support Vector Machine; DNN, Deep Neural Network; kNN, k-nearest Neighbor; K-ELM, Kernel-Based Extreme Learning Machine; LDA, Linear Discriminant Analysis; RF, Random Forest; GB, Gradient Boost; SOM, Self-Organizing Map; CNN, Convolutional Neural Network; LSTM, Long Short-Term Memory; DT, Decision Tree; ANN, Artificial Neural Network; cv, cross validation; loocv, leave-one-out cross validation.

approaches that try to reduce or completely negate the effects or causes of diseases and accidents.

4.1.1.1 Exposure Therapie

A very common type of therapy that leverages physiological data in virtual reality is exposure therapy. Heart rate, skin conductivity, or the respiration rate are often used to quantify anxiety reactions to stimuli that can be related to a phobia. Common examples for this are public speaking situations (Kothgassner et al., 2016; Kahlon et al., 2019), standing on elevated places (Gonzalez et al., 2016; Ramdhani et al., 2019), confrontations with spiders (Hildebrandt et al., 2016; Mertens et al., 2019), being locked up in a confined space (Shiban et al., 2016b; Tsai et al., 2018), or reliving a war-scenario (Almeida et al., 2016; Maples-Keller et al., 2019).

Physiological measures can also be used to evaluate the progress of the therapy. Shiban et al. (2017) created a virtual exposure application for the treatment of aviophobia. Heart rate and skin conductance were measured as indicators for the fear elicited by a virtual airplane flight. The exposure session consisted of three flights, while a follow-up test session, one week later, contained two flights. By analyzing the psychophysiological data throughout the five flights, the researchers were able to show that patients continuously got used to the fear stimulus.

Another way in which the physiological data can be utilized is for an automatic adaption of the exposure therapy system. Bălan et al. (2020) used a deep learning approach for the creation of a fear-level classifier based on heart rate, GSR, and EEG data. This classifier was then used as part of a virtual acrophobia therapy in which the immersed person stands on the roof of a building. Based on a target anxiety level and the output of the fear classifier the system can steer the height of the building and thus the intensity of the exposure. A similar approach comes from Herumurti et al. (2019) in the form of an exercise system for people with public speaking anxiety. Here, the behavior of a virtual audience depends on the heart rate of the user, i.e. the audience pays attention, pays no attention, or mocks the speaker.

What is also often done in research with exposure therapy applications is the comparison of different stimuli, systems, or groups of people. Physiological signals often represent a reference value that enhances such approaches. Comparisons have been made between traditional and virtual exposure therapy (Levy et al., 2016), fear-inducing stimuli in VR and AR (Li et al., 2017; Yeh et al., 2018), or just between phobic and healthy subjects (Breuninger et al., 2017; Kishimoto and Ding, 2019; Freire et al., 2020; Malta et al., 2020).

4.1.1.2 Relaxation Applications

Many approaches work with the idea to use a virtual environment to let people escape from their current situation and immerse

TABLE 3 | Sensors used in the identified classification approaches.

Manufacturer	Product	Measures	Type	More information	Description
BrainMaster	Freedom 24D Series	EEG	Headset	https://brainmaster.com/product/freedom-24-series/	Wireless EEG headset, 21 sensors
Tianyuan Xu, Ruixiang Yin, Lin Shu, Xiangmin Xu	Custom	Frontal EEG	HMD-integrated	Xu et al. (2019)	3 textile dry forehead electrodes, mounted inside the HMD
Pupil Labs	Binocular Add-on	ET	HMD-integrated	https://pupil-labs.com/products/vr-ar/	Bionocular eye-tracking camera for a HMD
Empatica	E4	PPG EDA SKT	Wristband	https://www.empatica.com/research/e4/	Wrist-worn device for real-time physiological data acquisition
Brain Products	actiCap	EEG	Electrode cap	https://www.brainproducts.com/products_by_type.php?tid=3	Various EEG caps with various channels and electrodes
Shimmer	Shimmer3 GSR+	PPG EDA	Wristband + Finger sensors	http://www.shimmersensing.com//products/gsr-optical-pulse-development-kit	Wrist placed unit that can be connected to finger-sensors via wires
tobii	TOBII VR	ET	HMD-integrated	https://vr.tobii.com	Bionocular eye-tracking camera for a HMD
Emotiv	Epoc+	EEG	Headset	https://www.emotiv.com/epoc/	14-channel EEG headset
Guger Technologies	g.GAMMACAP	EEG	Electrode cap	https://www.gtec.at/product/ggammasys/	16-channel electrode cap
G. Tartarisco et al.	PBS	ECG Respiration Motion	Chest band	Tartarisco et al. (2015)	Ergonomic chest band that integrates three sensors
interaxon	Muse 2	EEG PPG Motion Respiration	Headset	https://choosemuse.com/muse-2/	Wireless headband with 4 channels
emteq	FACETEQ	facial-EMG PPG Motion	HMD-integrated	https://www.emteqlabs.com/science/	Dry sensors, integrated into a VR headset
Thalmic Labs	Mayo	EMG	Armband	Discontinued	Wearable armband for EMG measurement
Biopac	B-Alert X-Series	EEG ECG	Head strip	https://www.biopac.com/product/b-alert-wireless-eeg-system/	9 or 20 channel wearable head strip that supports EEG and ECG
7invensun Technology	aGlass DKII	ET	HMD-integrated	https://www.7invensun.com/xrydxl	Bionocular eye-tracking camera for a HMD
Polar	Polar H10	ECG	Chest band	https://www.polar.com/us-en/products/accessories/h10_heart_rate_sensor	Wearable chest band for heart rate tracking
Biopac	PPG 100C	PPG	Amplifier	https://www.biopac.com/product/pulse-plethysmogram-amplifier/	Records blood volume pressure with a finger diode or an ear clip
Biopac	EDA 100C	EDA	Amplifier	https://www.biopac.com/product/eda-electrodermal-activity-amplifier/#product-tabs	Measures skin conductance with Ag/AgCl electrodes or electrode leads
antneuro	Waveguard toch	EEG	Electrode cap	https://www.ant-neuro.com/products/waveguard_touch	Cap for 8, 32, or 64 Ag/AgCl dry electrodes
Neuroscan	Quik-Cap	EEG	Electrode cap	https://compumedicsneuroscan.com/products/caps/quik-cap/	Cap available with 32, 64, or 128 channels

themselves in a more relaxing environment. Physiological stress indicators can help to assess the efficacy of these environments. Common examples for this include scenes with a forest (Yu et al., 2018; Browning et al., 2019; Wang X. et al., 2019; De Asis et al., 2020), a beach (Ahmaniemi et al., 2017; Anderson et al., 2017), mountains (Ahmaniemi et al., 2017; Zhu et al., 2019) or an underwater scenario (Soyka et al., 2016; Liszio et al., 2018; Fernandez et al., 2019). Other works go one step further and manipulate the virtual environment, based on the physiological status of the immersed person. So-called biofeedback applications are very common in the realm of relaxation applications and aim to make the users aware of their inner processes. The way this feedback looks can be very different. Blum et al. (2019) chose a virtual beach scene at sunset with palms, lamps, and a campfire. Their system calculates a real-time feedback parameter based on the heart rate variability as an indicator for relaxation. This parameter determines the cloud coverage in the sky and if the campfire and lamps are lit or not. Fominykh et al. (2018) present a similar virtual beach where the sea waves become higher and the clouds become darker when the heart rate of the user rises. Patibanda et al. (2017) present the serious game Life Tree which aims to teach a stress reducing breathing technique. The game revolves around a tree, that is bare at the start. By exhaling, the player can blow leaves towards the tree. The color of the leaves become green if the player breathes rhythmically and brown if not. Also, the color of the tree itself changes as the player practices correct breathing. Parenthoen et al. (2015) realized biofeedback with the help of EEG data by animating ocean

TABLE 4 | Various tools that were used in the identified classification approaches for the recording, synchronization, and processing of physiological data.

Tool	Reference	Type	Description	Example VR Application
SSI Framework	Wagner et al. (2013) Website: https://hcai.eu/projects/ssi/	Standalone framework	Synchronized processing of sensor data from multiple input devices and customizable machine learning pipeline. Supports recording, logging, annotating, processing, and pattern recognition.	Rangelova et al. (2019)
Pypsy	GitHub: https://github.com/brennon/Pypsy	Python library	Python library for processing and analyzing EDA data.	Saha et al. (2018)
EVE Framework	Grübel et al. (2016) GitHub: https://github.com/cog-ethz/EVE	Unity 3D plug-in	Unity based framework that facilitates the creation of custom VR experiments. Specific focus on sensor integration and data storage.	Weibel et al. (2018)
PhysioVR	Muñoz et al. (2016) GitHub: https://github.com/PhysioTools/PhysioVR	Unity 3D plug-in	Framework for the integration of physiological signals measured through wearable devices in mobile VR applications	Quintero et al. (2019)
NeuroRehabLab	Website: https://neurorehabilitation.m-iti.org/tools/en/	Tool collection	Website of the NeuroRehabLab research group with a collection of tools that help with the recording and processing of physiological data, e.g. demo project for connecting Emotiv Epoc with Unity.	
Mo-DBRS	Topalovic et al. (2020) GitHub: https://github.com/suthanalab/Mo-DBRS	Abstract platform	Platform that facilitates recording and synchronization of various physiological measurements and VR. Contains API for Unity, Python, and Matlab.	
Lab Stream Layer	GitHub: https://github.com/sccn/labstreaminglayer	Abstract platform	System that can handle networking, synchronization, access, and recording of measurements from various sources. Works with multiple platforms and languages.	Bălan et al. (2020)
VERA Project	Delvigne et al. (2020) GitHub: https://github.com/VDelv/VERA	Unity 3D project	Software that aims at facilitating attention related research in VR. The software also facilitates the recording of physiological signals.	
EEGLAB	Website: https://sccn.ucsd.edu/eeglab/index.php	Matlab toolbox	Toolbox for processing continuous and event-related EEG data (component analysis, artifact rejection, . . .)	Liu et al. (2014)
PREP Pipeline	Bigdely-Shamlo et al. (2015) GitHub: https://github.com/VisLab/EEG-Clean-Tools	Matlab toolbox	EEG processing pipeline that focuses on the identification of bad channels and the calculation of a robust average reference. It relies on the EEGLAB toolbox.	Hofmann et al. (2018)
Kurios	Tarvainen et al. (2014) Website: https://www.kubios.com	Standalone application	Heart rate variability analysis software. Supports various input formats and processing algorithms.	Blum et al. (2019)
EDA Explorer	Taylor et al. (2015) Website: https://eda-explorer.media.mit.edu	Python library	Collection of different scripts for processing EDA data.	Collins et al. (2019)
BioSPPy	Carreiras et al. (2015) https://github.com/PIA-Group/BioSPPy	Python library	Bundle of various signal processing and pattern recognition methods for physiological signals.	Salkevicius et al. (2019)
HeartPy	Van Gent et al. (2018) GitHub: https://github.com/paulvangentcom/heartrate_analysis_python	Python library	Toolbox for analyzing and processing heart rate data. Specialized towards noisy data.	Salkevicius et al. (2019)
BCI2000	Website: https://www.bci2000.org/mediawiki/index.php/Main_Page	Standalone application	Software for handling, recording, and analyzing EEG data.	Tremmel and Krusienski (2019)
MNE-Python	Gramfort et al. (2013) Website: https://mne.tools/stable/index.html	Python library	Python package for analyzing and visualizing neurophysiological data.	Delvigne et al. (2020)

waves according to surface cerebral electromagnetic waves of the immersed person. Most of these works aim at transferring the users from their stressful everyday life into a meditative state. Refer to Döllinger et al. (2021) for a systematic review of such works.

Relaxation applications can not only be used to escape the stress of everyday life but also as a distraction from painful medical procedures and conditions. This was applied in different contexts, e.g. during intravenous cannulation of cancer patients (Wong et al., 2020), preparation for knee surgery (Robertson et al., 2017), stay on an intensive care unit (Ong et al., 2020), or a dental extraction procedure (Koticha et al., 2019). Physiological stress indicators are commonly used to compare the effects of the virtual distraction to control groups (Ding et al., 2019; Hoxhallari et al., 2019; Rao et al., 2019).

4.1.1.3 Physical Therapy

VR stroke therapies often aim for the rehabilitation of impaired extremities. Here, virtual environments are commonly used to enhance motivation with gamification (Ma et al., 2018; Solanki and Lahiri, 2020) or to offer additional feedback, e.g. with a virtual mirror (Patel et al., 2015; Patel et al., 2017). In the domain of motor-rehabilitation, EMG-data can be of particular importance. It can be used to demonstrate the basic effectiveness of the system by showing that users of the application really activate targeted muscles (Park et al., 2016; Drolet et al., 2020). This is of special

TABLE 5 | Overview of the works from the field of therapy and rehabilitation that were discussed in **Section 4.1.1**. The Measures column refers to the physiological measures used in the work. The entries of the Independent Variables column often do not cover everything that was considered in the work. Entries in the Purposes column refer to the categories listed in **Section 4.1.6**.

Study	Scenario	Independent variables	Measures	Purposes
Kahlon et al. (2019)	Public speaking task	Before vs. during exposure	PPG	Process analysis
Kothgassner et al. (2016)	Public speaking task	Real audience vs. virtual audience vs. empty virtual hall	ECG Cortisol Secretion	Process analysis, stimuli comparison
Gonzalez et al. (2016)	Moving task on virtual platforms	Four different height levels	PPG, Motion	Stimuli comparison
Ramdhani et al. (2019)	Virtual ride in an open elevator	Different heights, pre-exposure vs. post-exposure	ECG, Resp, EDA	Process analysis, stimuli comparison
Hildebrandt et al. (2016)	Exposure to virtual spiders + collapsing floor + eerie sound	Cognitive flexibility of participants	EDA	Group comparison, correlation
Mertens et al. (2019)	Sitting on a virtual desk over which a spider walks	Spider fearful vs. no fearful participants	EDA, EMG	Group comparison, progress
Shiban et al. (2016b)	Sitting inside a virtual wooden box	Healthy vs. claustrophobic participants, visual cue vs. conceptual information vs. both	EDA, Resp, ECG PPG	Group comparison, stimuli comparison
Tsai et al. (2018)	Trapped in a virtual elevator during an emergency	Virtual Reality vs. Augmented Reality	ECG	Stimuli comparison
Maples-Keller et al. (2019)	Taking the perspective of a service member encountering a war scenario	Pre vs. Post PTSD exposure treatment, high vs. low responders	EMG, ECG, EDA	Progress, group comparison
Almeida et al. (2016)	Neutral and combat-related scenes	Veterans with PTSD vs. without PTSD, combat vs. classroom environment	ECG	Stimuli comparison, group comparison
Shiban et al. (2017)	Virtual airplane flight with subjects with aviophobia	With vs. without diaphragmatic breathing, first exposure vs. second exposure	ECG, EDA	Stimuli comparison, progress
Herumurti et al. (2019)	Public speaking task with a virtual audience that reacts to heart rate		PPG	Adaption
Levy et al. (2016)	Exposure to heights in a virtual skyscraper with acrophobic subjects	With vs. without physical presence of therapist	PPG	Progress, stimuli comparison
Li et al. (2017)	Doing a Stroop task in a virtual room with a sudden fire outbreak	Virtual Reality vs. Augmented Reality	ECG, EDA	Stimuli comparison, process analysis
Breuninger et al. (2017)	Sudden explosion in an underground garage	Healthy participants vs. patients with agoraphobia	ECG, EDA	Group comparison, process analysis
Kishimoto and Ding (2019)	Public speaking task with different types of audience	Virtual audience with ambiguous vs. negative feedback, healthy subjects vs. social anxiety patients	PPG	Group comparison, stimuli comparison
Freire et al. (2020)	Ride in a virtual, crowded bus	Healthy subjects vs. subjects with agoraphobia	ECG, EDA, Resp	Group comparison, process analysis
Malta et al. (2020)	Being under attack in a virtual war zone	Veterans with PTSD vs. without PTSD	ECG	Group comparison, progress
Wang X. et al. (2019)	Forest based resting environment	Seven different forest types	BP, PPG	Stimuli comparison
Yu et al. (2018)	3D videos of a crowded urban place and a forest environment	Forest vs. urban environment	PPG, BP Cortisol Secretion	Stimuli comparison, process analysis
Browning et al. (2019)	3D forest video	Real nature vs. VR nature video vs. indoor setting	EDA	Stimuli comparison
De Asis et al. (2020)	Visiting vacation spots in different settings	Before vs. during vs. after exposure, students with low vs. moderate vs. high stress	EDA, PPG	Group comparison
Ahmaniemi et al. (2017)	Visiting vacation spots in different settings during work	VR exposure vs. audio only	ECG, EDA, PPG, BP	Stimuli comparison, process analysis
Anderson et al. (2017)	Viewing different scenes after doing arithmetic test	Indoor vs. rural vs. beach scene	ECG, EDA	Stimuli comparison
Zhu et al. (2019)	Visiting vacation spots with different background music	Before vs. after exposure	EEG	Process analysis, correlation
Fernandez et al. (2019)	Underwater world with light flickering and sound pulsation		PPG, EEG	Feedback
Soyka et al. (2016)	Underwater world with a rhythmic moving jellyfish as a breathing guide	Baseline vs. underwater + jellyfish vs. only jellyfish	PPG, Resp	Stimuli comparison, correlation
Liszio et al. (2018)	Visiting a relaxing underwater world after stressful task	Desktop vs. HMD vs. no relaxation	ECG Cortisol Secretion	Stimuli comparison, process analysis
Blum et al. (2019)	Virtual beach that adapts to HRV of participants	VR vs. real relaxation session	ECG	Stimuli comparison, feedback
Fominykh et al. (2018)	Beach scene that adapts to the heart rate of the participant		PPG	Feedback

(Continued on following page)

TABLE 5 | (*Continued*) Overview of the works from the field of therapy and rehabilitation that were discussed in Section 4.1.1. The Measures column refers to the physiological measures used in the work. The entries of the Independent Variables column often do not cover everything that was considered in the work. Entries in the Purposes column refer to the categories listed in **Section 4.1.6.**

Study	Scenario	Independent variables	Measures	Purposes
Patibanda et al. (2017)	Virtual tree that changes its appearance according to breathing technique		Resp	Feedback
Parenthoen et al. (2015)	Flying over an animated ocean where waves are animated according to brain waves		EEG	Feedback
Robertson et al. (2017)	Exposure to a virtual beach before knee surgery	Standard hospital care vs. VR relaxation vs. tablet-based relaxation	BP, ECG, EDA	Stimuli comparison
Ong et al. (2020)	Exposure to calm beach scene with guided mediation during stay in an ICU	Individual relaxation sessions	ECG, Resp, BP	Progress
Koticha et al. (2019)	VR distraction for children undergoing extraction procedure	Dental extraction with vs. without VR	PPG	Stimuli comparison
Ding et al. (2019)	VR distraction during dressing change	Dressing change with vs. without VR	PPG	Stimuli comparison, process analysis
Hoxhallari et al. (2019)	Watching 3D nature video during tumescent local anesthesia	Anesthesia procedure with vs. without VR	PPG	Stimuli comparison, process analysis
Rao et al. (2019)	Viewing a 3D cartoon during restorative procedure	Baseline vs. during procedure vs. after procedure	PPG	Process analysis
Solanki and Lahiri (2020)	Walking on a virtual road during treadmill gait exercise	Healthy subjects vs. stroke subjects	ECG	Group comparison, process analysis
Ma et al. (2018)	Playing mini-games that require hand movement for stroke rehabilitation	Pre vs. mid vs. post rehabilitation	EMG	Feedback
Patel et al. (2015, 2017)	Doing hand activation task in front of a virtual mirror for stroke rehabilitation	Pre vs. post training, first day of training vs. last day	EMG	Progress
Park et al. (2016)	Lower extremity activation with feedback from a VR system for stroke rehabilitation	Slow vs. fast velocity VR training	EMG	Stimuli comparison
Vourvopoulos et al. (2019)	Hand activation for stroke patients in a virtual environment with different interaction types	EEG vs. EMG based motor feedback, pre intervention vs. during intervention vs. post intervention	EEG, EMG	Feedback, stimuli comparison, correlation, group comparison
Drolet et al. (2020)	Walking on a virtual walkway with changing underground while training on a treadmill.	Change of walkway underground realized through visual feedback vs. physical feedback vs. both	EMG	Stimuli comparison, process analysis
Dash et al. (2019)	VR basketball game that gives feedback about the strength of grip	Healthy subjects vs. stroke subjects, multiple exposures	EMG, EDA	Feedback, progress, group comparison
Calabro et al. (2017)	Walking in different virtual en-vironments during robot assisted gait training	Training with vs. without VR	EEG	Stimuli comparison, correlation
Ehgoetz Martens et al. (2015)	Walking across a virtual plank	Healthy subjects vs. subjects with Parkinson, ground-level plank vs. elevated plank	EDA	Stimuli comparison, group comparison
Ehgoetz Martens et al. (2016)	Standing on a virtual platform and walking on a virtual plank	Parkinson patients with vs. without gait impairment, walking vs. standing on a virtual plank	EDA	Stimuli comparison, group comparison
Gamito et al. (2014)	Exploring a virtual apartment with smoking cues (e.g. cigarettes, tobacco)	Environment with vs. without smoking cues, smokers vs. nonsmokers	ET	Stimuli comparison, group comparison
García-Rodríguez et al. (2013)	Exposure of smokers to a virtual pub	Smoking a cigarette in the pub vs. playing darts in the pub vs. freely exploring the pub	ECG	Stimuli comparison, process analysis
Thompson-Lake et al. (2015)	Sitting in a virtual room with smoking related and non-smoking related cues	Smoking cues vs. non-smoking cues	PPG	Stimuli comparison
Yong-Guang et al. (2018)	Video with METH-cues	Subjects with vs. without METH dependence	ECG	Group comparison
Ding X. et al. (2020)	Neutral environment and environ-ment with avatars using METH	Subjects with vs. without METH dependence, environment with vs. without METH cues	EDA, EEG	Stimuli comparison, group comparison
Wang Y.-G. et al. (2019)	Watching 3D videos that show negative consequences of METH	People with vs. without counter-conditioning therapy	ECG	Group comparison
Bruder and Peters (2020)	Exploring a virtual casino, a sports betting facility and a cafe	Environment with vs. without gambling cues	EDA	Stimuli comparison, process analysis
Detez et al. (2019)	Exploring a VR casino-bar with slot machines	Wins vs. loses vs. near-misses on slot machine	ECG	Stimuli comparison, process analysis

interest when impairments do not allow visible movement of the target-limb (Patel et al., 2015). Another strategy to combine VR stroke therapy and EMG signals is a feedback approach. Here, the strength of the muscle activation is made available to the user visually or audibly which can result in positive therapy effects, as the user becomes aware of internal processes (Dash et al., 2019;

Vourvopoulos et al., 2019). The use of psychophysiological data in stroke-therapy is not necessarily restricted to EMG. Calabro et al. (2017) created a virtual gait training for lower limb paralysis and compared it to a non-VR version of the therapy. With the help of EEG measurements, they showed that the VR version was especially useful for activating brain areas that are responsible for motor learning.

Also Parkinson's disease requires motor-rehabilitation. Researchers have used physiological data to assess the anxiety experience of Parkinson patients with impaired gait under different elevations or on a virtual plank (Ehgoetz Martens et al., 2015; Ehgoetz Martens et al., 2016; Kaur et al., 2019). This data has helped researchers and therapists to understand the experience of the patients and to adapt the therapy accordingly.

4.1.1.4 Addiction Therapy

Another field of application where physiological measurements prove useful is in the therapy of drug addictions. Gamito et al. (2014) showed that virtual cues have the potential to elicit a craving for nicotine in smokers. With the help of eye-tracking, they demonstrated that smokers exhibit a significantly higher number of eye fixations on cigarettes and tobacco packages. In a similar studies, Thompson-Lake et al. (2015) and García-Rodríguez et al. (2013) showed that virtual, smoking-related cues can cause an increase in the heart rate of addicts. Yong-Guang et al. (2018) and Ding X. et al. (2020) did the same for methamphetamine users. They found evidence that meth-users show significant differences in EEG, GSR, and heart rate variability measurements when being exposed to drug-related stimuli in a virtual environment. Based on this, Wang Y.-G. et al. (2019) created a VR counter-conditioning procedure for methamphetamine users. With this virtual therapy, they were able to suppress cue-induced reactions in patients with meth-dependence. The use of physiological data to study the effects of addiction-related stimuli that are presented in VR has also been applied for gambling (Bruder and Peters, 2020; Detez et al., 2019).

A summary of the works discussed in this section can be found in **Table 5**.

4.1.2 Training and Education

A considerable amount of VR applications help people to learn new skills, enhance existing ones, or facilitate knowledge in a certain area. A major reason why physiological data comes in handy in training and teaching applications is its potential to indicate cognitive workload and the stress state of a human subject.

4.1.2.1 Simulator Training

Training simulators from various domains include an estimation of mental workload based on physiological data, e.g. surgery training (Gao et al., 2019), virtual driving (Bozkir et al., 2019), or flight simulation (Zhang S. et al., 2017). One way to use knowledge about cognitive load is by adapting task difficulty. Dey et al. (2019a) created a VR training task that requires the user to select a target object, defined by a combination of shape and color. The system uses the EEG alpha band signal to determine how demanding the task is. Based on this information the system can steer the difficulty of the task by altering the number and properties of distractors. In this way, it can be ensured that the task is neither too easy nor too difficult. In the application of Faller et al. (2019) the user has to navigate a plane through rings. The difficulty, i.e. the size of the rings, can be adjusted based on EEG data. With this approach, they were able to keep trainees on an arousal level that is ideal for learning.

In certain fields, physiological measures can even be used to determine the difference between experts and novices. Clifford et al. (2018) worked with a VR application for the training of aerial firefighting. It is a multi-user system that requires communication from the trainees. To cause additional stress the system includes a scenario where the communication is distorted. They evaluated the system with novice and experienced firefighters. By analyzing the heart rate variability of subjects, they were able to show that the communication disorders were effective in eliciting stress throughout the subjects. More interestingly, however, experts showed an increased ability to maintain their heart rate variability, compared to the inexperienced firefighters. This indicates that they were better able to cope with the stress (Clifford et al., 2020). Currie et al. (2019) worked with a similar approach. Their virtual training environment is focused on a high-fidelity surgical procedure. Eye-tracking was used to gain information about the attention patterns of users. A study with novice and expert operators showed that the expert group had significantly greater dwell time and fixations on support displays (screens with X-ray or vital signs). Melnyk et al. (2021) showed how this knowledge can be used to support learning, as they augmented surgical training by using expert gaze patterns to guide the trainees.

In simpler cases, stress and workload indicators are used to substantiate the basic effectiveness of virtual stimuli in training. Physiological indicators can be used to show that implemented scenarios are really able to elicit desired stress responses (Loreto et al., 2018; Prachyabrued et al., 2019; Spangler et al., 2020).

4.1.2.2 Virtual Classrooms

Physiological measures can also benefit classical teacher-student scenarios. Rahman et al. (2020) present a virtual education environment in which the teacher is provided with a visual representation of the gaze behavior of students. This allows the teachers to identify distracted or confused students, which can benefit the transfer of knowledge. Yoshimura et al. (2019) developed a strategy to deal with inattentive listeners. They constructed an educational virtual environment in which eye-tracking is used to identify distracted students. The system can then present visual cues, e.g. arrows or lines, that direct the attention of the pupils towards critical objects that are currently discussed. In the educational environment of (Khokhar et al., 2019) the knowledge about inattentive students is provided to a pedagogical agent. Sakamoto et al. (2020) tested pupil metrics for their eligibility to gain information about the comprehension of people. They recorded gaze behavior during a learning task in VR and compared this to the subjective comprehension ratings of subjects. In a similar example from Orlosky et al. (2019), they used eye movement and pupil size data to build a support vector

TABLE 6 | Overview of the works from the field of training and education that were discussed in **Section 4.1.2**. The Measures column refers to the physiological measures used in the work. The entries of the Independent Variables column often do not cover everything that was considered in the work. Entries in the Purposes column refer to the categories listed in **Section 4.1.2**.

Study	Scenario	Independent variables	Measures	Purposes
Gao et al. (2019)	Virtual reality laparoscopic surgery simulator with a secondary arithmetic task	Double task vs. single task, experienced vs. inexperienced surgeons	ET	Stimuli comparison, group comparison
Zhang S. et al. (2017)	Virtual flight simulator	3D vs. 2D simulation, task difficulty	EEG	Stimuli comparison
Bozkir et al. (2019)	Virtual driving simulator	Normal driving vs. situation with critically crossing pedestrian	ET	Classification
Dey et al. (2019a)	Virtual object selection task with different levels of difficulty that are adapted to information from EEG	Task difficulty	EEG	Adaption, stimuli comparison
Faller et al. (2019)	Navigate a plane through a course of rings, where the difficulty can be adapted to feedback from EEG	EEG Feedback vs. sham feedback vs. no feedback	ET, ECG EEG	Adaption, classification stimuli comparison
Clifford et al. (2018), Clifford et al. (2020)	Aerial firefighting simulation with special communication requirements	Situations with. vs. without communication disruptions, experts vs. novices	ECG Resp	Stimuli comparison, group comparison
Currie et al. (2019)	Virtual coronary angiography procedure with a secondary card game	Novice operators vs. expert operators	ET, EDA, motion, PPG	Group comparison
Melnyk et al. (2021)	Virtual surgery training with secondary counting task	Training with gaze feedback vs. training with movement feedback	ET	Feedback, stimuli comparison
Prachyabrued et al. (2019)	Playing as an emergency worker on a stressful rescue mission	Emotional connection with virtual co-worker vs. no emotional connection, baseline vs. training	PPG, EDA	Stimuli comparison
Spangler et al. (2020)	Virtual shooting tasks with different levels of difficulty	High vs. low difficulty tasks, number of sessions	ECG, BP	Stimuli comparison, progress
Loreto et al. (2018)	Virtual reality work-at-height simulation while climbing a ladder in real-life	Simulation with vs. without vibration feedback	EDA	Stimuli comparison
Rahman et al. (2020)	Virtual education environment based on a solar field + visualization of gaze	Different gaze visualization techniques, single user vs. multiuser VR	ET	Feedback
Yoshimura et al. (2019)	Virtual educational environment involving an oil rig + pedagogical agent + visual cues for attention restoration	Different kinds of virtual cues that mark the current point of interest	ET	Feedback
Khokhar et al. (2019)	Virtual educational environment involving an oil rig + pedagogical agent that is sensible towards attention shifting		ET	Feedback
Sakamoto et al. (2020)	Virtual environment with comic-based educational material + subjective rating of comprehension		ET	Correlation
Bian et al. (2016)	VR shooting game + subjective rating of flow		ECG, Resp EMG, EDA	Correlation
Mishra and Folmer (2018)	Virtual exercise game based on the collection of objects	Pre vs. post exercise	PPG	Process analysis
Kivelä et al. (2019)	Two different virtual exercise games	BeatSaber vs. QuiVR	PPG	Stimuli comparison
Debska et al. (2019)	Virtual obstacle course on an omni tread-mill and a flight simulator where steerage works with body movement	Omni treadmill vs. flight simulator	PPG	Stimuli comparison
Zeng et al. (2017)	Playing arcade mini-games where loco-motion is realized with an exercise bike.	Traditional vs. VR-based bike exercise	BP	Stimuli comparison
McDonough et al. (2020)	Playing arcade mini-games where loco-motion is realized by a exercise bike.	Traditional vs. VR-based vs. desktop-based bike exercise	BP	Stimuli comparison
Campbell and Fraser (2019)	Virtual cycling game in which the speed depends on the user's heart rate which is also displayed in the HUD	HMD vs. 2D screen, physical resistance, true vs. falsified heart rate feedback	PPG	Feedback, Adaption
Kirsch et al. (2019)	Bow and arrow-based high intensity interval training which can adapt to the user's heart rate	Adaption of music vs. adaption of lighting based on heart rate	PPG	Adaption
Yoo et al. (2018)	VR exercise game in which the player must throw snowballs at waves of enemies and in which heart rate is displayed		PPG	Feedback
Kojić et al. (2019)	VR rowing game with feedback about the respiration rate	Different feedback methods vs. no feedback vs. no VR	Resp	Feedback, stimuli comparison
Greinacher et al. (2020)	Virtual rowing exercise with the aim to promote breathing-movement-synchrony	Verbal vs. visual vs. tactile respiration feedback	Resp	Feedback, stimuli comparison

machine that predicts if a user understood a given term or not. Even the experience of flow can be assessed with the help of physiological indicators (Bian et al., 2016). Information about attention and comprehension of students can be used to optimize teaching scenarios. It is an illustrative example of how physiological data can augment virtual learning spaces and create possibilities that would be unthinkable in real-life ones.

4.1.2.3 Physical Training

Our discussion of training applications has thus far revolved around mental training. However, there are also applications for physical training in VR. Again, physiological data can be used to emphasize the basic effectiveness of the application. Changes in heart rate or oxygen consumption can show that virtual training elicits physical exertion and give insights into the extent of it (Mishra and Folmer, 2018; Xie et al., 2018; Debska et al., 2019; Kivelä et al., 2019). This can also provide a reference value for the comparison of real-life and virtual exercising. Works like Zeng et al. (2017) and McDonough et al. (2020) compared a VR-based bike exercise with a traditional one. Their assessment of exertion with the help of BP measurements showed no significant difference between the virtual and analog exercises. Measurements of the subjectively perceived exertion, however, showed that participants of a VR-based training felt significantly less physiological fatigue.

As in other fields of application, physiological data can be used to adapt the virtual environment. Campbell and Fraser (2019) present an application where the trainee rides a stationary bike while wearing an HMD. In the virtual environment, the user is represented by a cyclist avatar. The goal of the training is to cover as much distance as possible in the virtual world, however, the speed at which the avatar moves is determined by the heart rate of the user. This way, the difficulty cannot be reduced by simply reducing the resistance of the bike and unfit users have the opportunity to cover more distance. In the exercise environment of Kirsch et al. (2019), the music-tempo is adapted to the user's heart rate which was perceived as motivating by the trainees. Other works just take the physiological data and display it to the users so they can keep track of their real-time physical exertion (Yoo et al., 2018; Kojić et al., 2019; Greinacher et al., 2020).

A summary of the works discussed in this section can be found in **Table 6**.

4.1.3 Entertainment

Another field of application comprises VR systems that are primarily built for entertainment purposes, i.e. games and videos. Often, researchers use physiological data to get information about the arousal video or a game elicits (Shumailov and Gunes, 2017; Ding et al., 2018; Mavridou et al., 2018b; Ishaque et al., 2020). Physiological measures can also be used as explicit game features. For example, progress may be denied if a player is unable to adjust their heart rate to a certain level (Houzangbe et al., 2019; Mosquera et al., 2019). Additionally, the field of view in a horror game can be adjusted depending on the heartbeat (Houzangbe et al., 2018). Kocur et al. (2020) present a novel way to help novice users in a shooter game by introducing a gaze-based aiming assistant. If the user does not hit a target with his shot, the assistant can guess what the actual target was, based on the gaze. When the shot is close enough to the intended target it hits nevertheless. Moreover, eye-tracking can be used to optimize VR video streaming. Yang et al. (2019) used gaze-tracking to analyze the user's attention and leverage this knowledge to reduce the bandwidth of video streaming by reducing the quality of those parts of a scene that are not focused.

A summary of the works discussed in this section can be found in **Table 7**.

4.1.4 Functional Virtual Reality Properties

Within this section, we describe applications that make use of common techniques of VR. These are applications that include embodiment, agent interaction, or multiuser VR. We are talking about functional properties that may or may not be part of the system. These properties can also be part of the fields of application that we discussed before, e.g. therapy and training. Nevertheless, we have identified them as separate fields because physiological measurements have their own functions in applications that use embodiment, agent interaction, or multiuser VR. Researchers who want to use those techniques in their own applications can find separate information about the role of the physiological measurements here.

4.1.4.1 Applications With Embodiment

A range of VR applications use avatars as a representation of the user in the virtual realm (Lugrin et al., 2018; Lugrin et al., 2019b; Wolf et al., 2020). VR has the potential to elicit the illusion of owning a digital body which can be referred to as the Illusion of Virtual Body Ownership (Lugrin et al., 2015; Roth et al., 2017). This concept is an extension of the rubber-hand illusion (Botvinick and Cohen, 1998), which has the consequence that the feeling of ownership is often based on the synchrony of multi sensory information, e.g. visuo-tactile or visuo-motor (Tsakiris et al., 2006; Slater et al., 2008). Physiological data can provide information about whether and to what extent the virtual body is perceived as the own. One way to provide objective evidence for the illusion of body ownership is to threaten the artificial body-part while measuring the skin response to get information about whether the person shows an anxiety reaction (Armel and Ramachandran, 2003; Ehrsson et al., 2007). One of the most common paradigms still used in more recent literature is to threaten the virtual body (part) with a knife stab (González-Franco et al., 2014; Ma and Hommel, 2015; Preuss and Ehrsson, 2019). Alchalabi et al. (2019) present an approach that uses EEG data to estimate embodiment. They worked with a conflict between visual feedback and motor control. That means subjects had to perform a moving task on a treadmill that was replicated by their virtual representation. However, the avatar stopped walking prematurely while the subject was still moving in real life. This modification in feedback was reflected in EEG data and results showed a strong correlation between the subjective level of embodiment and brain activation over the motor- and pre-motor cortex. Relations between EEG patterns and the illusion of body ownership were also shown in virtual variations of the rubber-hand illusion (González-Franco et al., 2014; Skola and Liarokapis, 2016). Furthermore, there is also evidence that the feeling of ownership and agency over a virtual body or limb can be reflected in skin temperature regulation (Macauda et al., 2015; Tieri et al., 2017).

Other works connect embodiment and physiological measures by investigating how the behavior or properties of an avatar can change physiological responses. In their study, Czub and Kowal (2019) introduced a visuo-respiratory conflict, i.e. the avatar that

TABLE 7 | Overview of the works from the field of entertainment that were discussed in **Section 4.1.3**. The Measures column refers to the physiological measures used in the work. The entries of the Independent Variables column often do not cover everything that was considered in the work. Entries in the Purposes column refer to the categories listed in **Section 4.1.6**.

Study	Scenario	Independent variables	Measures	Purposes
Ding et al. (2018)	Watching clips from the movie The Jungle Book	Traditional 2D film vs. VR film	SKT, ECG, Resp, PPG	Stimuli comparison
Mosquera et al. (2019)	Puzzle game on a virtual spaceship that offers HR feedback and requires control over the HR		PPG	Feedback, adaption, process analysis
Houzangbe et al. (2019)	Puzzle game that requires control over the HR	Subjects with different levels of HR control	PPG	Feedback, adaption, classification
Xie et al. (2018)	VR exercise games with procedural level design	Easy vs. medium vs. hard physical difficulty	ECG	Stimuli comparison
Houzangbe et al. (2018)	VR horror game in which field of view and sound is adapted to the heart beat	Game with vs. without adaption mechanics	PPG	Feedback, adaption
Kocur et al. (2020)	VR first-person shooter with a gaze-based aiming assistant	No aiming assistance vs. standard assistance vs. gaze-based assistance	ET	Adaption
Yang et al. (2019)	VR video streaming in which quality of the non-focused parts is reduced	Different kinds of movies	ET	Adaption

represented the subjects showed a different respiration rate than its owner. They found out that the immersed subjects actually adapted their respiration rate to their virtual representation. The frequency of breathing increased when the breathing animation of the avatar was played faster and vice versa. Kokkinara et al. (2016) showed that activity of the virtual body, i.e. climbing a hill, can increase the heart rate of subjects, even if they are sitting on a chair in real life.

Another link between physiological measures and body-ownership can be made when those measures are used as input for the behavior of the avatar. Betka et al. (2020) executed a study in which they measured the respiration rate of the subjects and mapped it onto the avatar that was used as their virtual representation. Results showed that congruency of breathing behavior is an important factor for the sense of agency and the sense of ownership over the virtual body.

4.1.4.2 Applications With Agent Interaction

In the last section, we focused on applications that use avatars to embody users in the virtual environment. Now we move from the virtual representation of the user to the virtual representation of an artificial intelligence, so-called agents (Luck and Aylett, 2000).

Physiological data is often used to analyze and understand the interaction between a human user and agents. The study of Gupta et al. (2019), that was revised in Gupta et al. (2020), aimed to learn about the trust between humans and agents. The primary task of this study comprised a shape selection where subjects had to find a target object that was defined by shape and color. An agent was implemented that gave hints about the direction in which the object could be found. There were two versions of the agent, whereat one version always gave an accurate hint and the other one did not. With the help of a secondary task, an additional workload was induced. EEG, GSR, and heart rate variability were captured throughout the experiment, as an objective indicator for the cognitive workload of the subjects. In the EEG data, Gupta et al. (2020) found a significant main effect for the accuracy of the agent's hints. That means subjects who received correct hints showed less cognitive load. The authors interpret this as a sign of trust towards the agent as the subjects did not seem to put any additional effort into the shape selection task as soon as they got the correct hints from the virtual assistant. In another example, Krogmeier et al. (2019) investigated the effects of bumping into a virtual character. In their study, they manipulated the haptic feedback during the collision. They explored how this encounter and the introduction of haptic feedback changed the physiological arousal of the subjects gauged with EDA. In a related study, Swidrak and Pochwatko (2019) showed a heart-rate deceleration of people who are touched by a virtual human. Another facet of human-agent interaction is the role of different facial expressions and how they affect physiological responses (Mueller et al., 2017; Ravaja et al., 2018; Kaminskas and Sciglinskas, 2019).

Other works investigate different kinds of agents and use physiological data as a reference for their comparison. Volante et al. (2016) investigated different styles of virtual humans, i.e. visually realistic vs. cartoon-like vs. sketch-like. The agents were depicted as patients in a hospital which showed progressive deterioration of their medical condition. With the help of EDA data, they were able to quantify emotional responses towards those avatars and analyze how these responses were affected by the visual appearance. Other works compared gaze behavior during contact with real people and agents (Syrjamaki et al., 2020) or the responses to virtual crowds showing different emotions (Volonte et al., 2020).

Another category can be seen in studies that leverage agents to simulate certain scenarios and use physiological data to test the efficacy of these scenarios to elicit desired emotional responses. At this point, there is a relatively large overlap with the previously discussed exposure therapies (**Section 4.1.1.1**). Applications aimed at the treatment of social anxiety often include the exposition to a virtual audience that aims to generate a certain atmosphere (Herumurti et al., 2019; Lugrin et al., 2019a; Streck et al., 2019). Kothgassner et al. (2016) asked participants of their study to speak in front of a real and a virtual audience. Heart rate, heart rate variability, and saliva cortisol secretion were assessed. For both groups, these stress indicators increased similarly, which demonstrates the fundamental usefulness of such therapy systems, as the physiological response to a virtual audience was comparable to a real one. Other studies investigated stress reactions depending on the size (Mostajeran et al., 2020) or displayed emotions (Barreda-Angeles et al., 2020) of the audience. The potential of virtual audiences to elicit stress is not only applicable to people with social anxiety. Research approaches that investigate human behavior and experience under stress can use a speech task in front of a virtual audience as a stressor. This can be referred to as the Trier Social Stress Test, which was often transferred to the virtual realm (Delahaye et al., 2015; Shiban et al., 2016a; Kothgassner et al., 2019; Zimmer et al., 2019; Kerous et al., 2020). Social training applications that work with virtual audiences are also available specifically for people with autism. Again, physiological measurements help to understand the condition of the user and thus to adjust the training (Kuriakose et al., 2013; Bekele et al., 2016; Simões et al., 2018). Also physical training applications can use physiological data to determine the effect of agents. Murray et al. (2016) worked with a virtual aerobic exercise, i.e. rowing on an ergometer. One cohort of their study had a virtual companion that performed the exercise alongside the subject. In a related study, Haller et al. (2019) investigated the effect of a clapping virtual audience on the performance in a high-intensity interval training. In both examples the effect of the agents was evaluated with a comparison of the heart rate. It indicates changes in the physical effort and can thus show whether the presence of agents changes training behavior.

4.1.4.3 Applications With Multiuser Virtual Reality

In multiuser VR applications, two or more users can be present and interact with each other at the same time (Schroeder, 2010). This concept offers the possibility of exchanging physiological data among those users. Dey et al. (2018) designed three different collaborative virtual environments comprising puzzles that must be solved together. They evaluated those environments in a user-study, whereat one group got auditory and haptic feedback about the heart rate of the partner. Results indicated that participants who received the feedback felt the presence of the collaborator more. There is even evidence that the heart rate feedback received from a partner can cause an adaption in the own heart rate (Dey et al., 2019b). In a similar approach, Salminen et al. (2019) used an application that shares EEG and respiration information among subjects in a virtual meditation exercise. The feedback was depicted as a glowing aura that pulsates according to the respiration rate and is visualized with different colors, depending on brain activity. Users who had this kind of feedback perceived more empathy towards the other user. Desnoyers-Stewart et al. (2019) built an application that deliberately aims to achieve such synchronization of physiological signals in order to establish a connection between users. Another way in which multiuser VR applications can benefit from physiological measures is in terms of communication. Lou et al. (2020) present a hardware solution that uses EMG sensors to track facial muscle activity. These activities are then translated to a set of facial expressions that can be displayed by an avatar. This offers the possibility of adding nonverbal cues to interpersonal communication in VR.

A summary of the works discussed in this section can be found in **Table 8**.

4.1.5 General Virtual Reality Properties

Our last field of application focuses on properties that are relevant for every VR application as they are inherent to the medium itself. These are cybersickness and presence. Here we are talking about non-functional properties of a VR system, as they can occur to varying degrees. These varying degrees of cybersickness and presence are either actively manipulated or passively observed. In both cases, consideration of physiological measurements can provide interesting insights.

4.1.5.1 Presence

Presence describes the experience of a user to be situated in the virtual instead of the real environment Witmer and Singer (1998). Hence, knowledge about the extent to which a virtual environment can elicit the feeling of presence in a user is relevant in most VR applications. Beyond the classic presence questionnaires from Slater et al. (1994) or Witmer and Singer (1998), there are also approaches that aim to determine presence based on physiological data.

Athif et al. (2020) present a comprehensive study that relates presence factors to physiological signals. They worked with a VR forest scenario in which the player needs to collect mushrooms that spawn randomly. This scenario was implemented in six different gradations based on the four factors of presence, described by Witmer and Singer (1998). These are distraction, control, sensory, and realism. That means the base version fulfilled the requirements for all these factors. Four versions suppressed one factor each and one version suppressed all the factors simultaneously. In the study, participants were presented with each of the scenarios while their physiological reactions were measured. Data showed that EEG features indicated changes in presence particularly well, while ECG and EDA features did not. Signals from temporal and parietal regions of the brain showed correlations with the suppression of the specific presence factors. In a similar investigation Dey et al. (2020) implemented two versions of a cart-ride through a virtual jungle. Their high presence version was realized through higher visual fidelity, more control, and object-specific sound. In this setup, they were able to show a significant increase in the heart rate of people presented to the high presence version, whereat EDA showed no systematic changes. The study of Deniaud et al. (2015) showed correlations between presence questionnaire scores, skin conductance, and heart rate variability. Other studies again, found the heart rate or EDA data to be weak indicators for presence (Felnhofer et al., 2014; Felnhofer et al., 2015). Szczurowski and Smith (2017) suggest to gauge presence through a comparison of virtual and real stimuli. Accordingly, a high presence is characterized in such a way that the exposure to the virtual stimulus elicits similar physiological responses as the exposure to the real stimulus. As such, one could take any physiological measure to gauge presence, as long as one has a comparative value from a real life stimulus.

The exact relationship between specific physiological measures and the experience of presence still seems ambiguous. This may also be due to the fact that the concept of presence is understood

TABLE 8 | Overview of the works from the field of functional VR properties that were discussed in **Section 4.1.4**. The Measures column refers to the physiological measures used in the work. The entries of the Independent Variables column often do not cover everything that was considered in the work. Entries in the Purposes column refer to the categories listed in **Section 4.1.6**.

Study	Scenario	Independent variables	Measures	Purposes
González-Franco et al. (2014)	Sitting on a virtual table while being embodied in a first-person perspective	Virtual knife attacking the hand vs. attacking the table, actively moving hand vs. no movement	EEG	Stimuli comparison, correlation, process analysis
Ma and Hommel (2015)	Virtual rubber hand illusion with a knife attack on virtual body parts	Synchronous vs. asynchronous vibro-tactile simulation, embodiment through hand vs. embodiment through rectangle	EDA	Stimuli comparison
Preuss and Ehrsson (2019)	Scenario in which body ownership is induced with galvanic vestibular stimulation and a virtual knife attack	Synchronous vs. asynchronous visuo-vestibular stimulation	EDA	Stimuli comparison
Alchalabi et al. (2019)	Walking through a virtual corridor while being on a treadmill in real life	Perform vs. watch vs. imagine walking, synchronous vs. asynchronous movement	EEG	Stimuli comparison, correlation
Skola and Liarokapis (2016)	Rubber hand illusion	VR vs. AR vs. real life	EEG	Stimuli comparison, correlation
Tieri et al. (2017)	Sitting on a virtual table with embodiment of the arms	Limb embodiment through a hand vs. hand, detached from arm vs. wood block, observing virtual limb vs. observing a ball	ECG, SKT	Stimuli comparison, correlation
Macauda et al. (2015)	Watching a 3D video while being on a motion platform and being embodied in first person perspective	Visuo-vestibular synchronization vs. delay, embodiment through mannequin vs. red pillow	SKT	Stimuli comparison, correlation
Czub and Kowal (2019)	Sitting on a virtual bench while being embodied in first person perspective with an avatar that depicts breathing motion	Visuo-respiratory synchronization vs.no synchronization	Resp	Feedback, stimuli comparison
Kokkinara et al. (2016)	Embodying a virtual avatar that is climbing a hill while sitting on a stool in real life	First person perspective vs. third person perspective, sway animation vs. no sway animation	ECG, Resp, EDA	Stimuli comparison, correlation, process analysis
Betka et al. (2020)	Watching an avatar with a flashing outline from third person perspective	Synchrony vs. asynchrony between respiration and flashing of avatar, active vs. passive breathing	Resp	Feedback, stimuli comparison
Gupta et al. (2019), Gupta et al. 2020)	Object search task with a virtual voice agent that gives indications regarding the target	High vs. low difficulty search task, high vs. low accuracy indications	EEG, EDA PPG	Stimuli comparison, correlation
Krogmeier et al. (2019)	Scene in which numerous virtual agents walk past or collide with the user	Different kinds of haptic feedback vs. no haptic feedback when colliding with agents	EDA	Stimuli comparison, correlation
Swidrak and Pochwatko (2019)	Playing a decisions-making game with a virtual agent that touches the subject during the procedure	Gender, stereotypical femininity /masculinity, apparent social status of agent, touch with no vs. acoustic vs. tactile feedback	ECG	Stimuli comparison
Ravaja et al. (2018)	Playing a prisoner's dilemma game with a virtual agent	Facial expression of the agent	EMG, EEG	Stimuli comparison adaption, correlation
Mueller et al. (2017)	Sitting on a virtual table, facing an agent when a sudden noise burst appears	Violet vs. teal room, 95 vs. 80 db noise burst, neutral vs. angry facial expression of the agent	EEG	Process analysis, stimuli comparison
Volante et al. (2016)	VR training system to help nurses identify the signs of rapid patient deterioration	Visually realistic vs. cartoon-like vs. sketch-like patient	EDA	Stimuli comparison
Syrjamaki et al. (2020)	Face-to-face situation with a virtual avatar or a real person	VR vs. face-to-face interaction, direct vs. averted gaze	EDA, ECG	Stimuli comparison
Volonte et al. (2020)	Virtual market simulation in which the subject has to get items from different vendors	Virtual crowd with positive vs. negative vs. neutral vs. mixed emotional expressions	EDA	Stimuli comparison
Streck et al. (2019)	Different virtual environments that contain virtual crowds, e.g. classroom, library, bar		EDA, ECG ET	Feedback
Herumurti et al. (2019)	Public speaking task in front of a virtual audience whose behavior is adjusted depending on the heart rate.		PPG	Adaption
Mostajeran et al. (2020)	Giving a speech in front of a virtual audience and performing arithmetic calculations (Trier Social Stress Test)	Three vs. five vs. fifteen agents in virtual audience vs. real audience with three people	Salivary cortisol, ECG, EDA	Stimuli comparison, correlation
Barreda-Angeles et al. (2020)	Public speaking task in front of a 360°-video audience	Neutral vs. positive vs. negative reaction of the virtual audience	ECG, EDA	Stimuli comparison, group comparison
Kothgassner et al. (2019)	Giving a speech in front of a virtual audience and performing arithmetic calculations with prior social support	Real vs. avatar-based vs. agent-based vs. no social support	PPG	Stimuli-Comparison, process analysis
Shiban et al. (2016a)	Giving a speech in front of a virtual audience and performing arithmetic calculations	Doing the task in real life vs. VR vs. VR with a virtual competitor	Salivary cortisol, ECG, EDA	Stimuli comparison, process analysis

(Continued on following page)

TABLE 8 | (*Continued*) Overview of the works from the field of functional VR properties that were discussed in Section 4.1.4. The Measures column refers to the physiological measures used in the work. The entries of the Independent Variables column often do not cover everything that was considered in the work. Entries in the Purposes column refer to the categories listed in **Section 4.1.6.**

Study	Scenario	Independent variables	Measures	Purposes
Zimmer et al. (2019)	Giving a speech in front of a virtual audience and performing arithmetic calculations	Speak in front of a virtual vs. real vs. no audience	Salivary cortisol, PPG, EDA	Stimuli comparison, process analysis
Delahaye et al. (2015)	Giving a speech in front of a virtual audience and navigation through two labyrinths	Speech task vs. labyrinth task	ECG	Stimuli comparison, process analysis, correlation
Kerous et al. (2020)	Doing a virtual Stroop task while being observed by virtual agents as social stressors	Only social stressor vs. only Stroop vs. combination of both	ECG, EDA	Stimuli comparison
Bekele et al. (2016)	Virtual agent based communication training for people with autism	Agent with vs. without gaze sensitivity, different sessions	ET, ECG, EDA SKT	Progress, adaption stimuli comparison
Simões et al. (2018)	Serious game for teaching people with autism to get used to bus-taking routines, especially the social situations	People with vs. without autism, different sessions	EDA	Group comparison, progress
Kuriakose et al. (2013)	Different encounters with virtual agents for training proper reactions in specific social situations for people with autism	Difficulty levels of the social situations	PPG, SKT	Stimuli comparison, progress analysis, correlation
Murray et al. (2016)	Sitting in a virtual rowing boat with a virtual agent while training with a rowing ergometer in real life	Rowing vs. rowing in VR vs. rowing in VR with agent support	ECG	Stimuli comparison, process analysis, correlation
Haller et al. (2019)	Playing a virtual bike racing game while training on an ergometer in real-life	Exercise with vs. without supporting virtual crowd	PPG	Stimuli comparison
Dey et al. (2018)	Solving virtual escape room puzzles together with a partner	With vs. without audio-haptic heart rate feedback of the partner	ECG	Feedback
Dey et al. (2019b)	Active (shooter) and passive (safari) multiuser virtual environments	Decreased vs. unchanged vs. increased heart rate feedback of the partner, active vs. passive virtual environment	ECG	Feedback, stimuli comparison
Salminen et al. (2019)	Virtual multiuser meditative environment with feedback about the own and the partner's EEG and respiration signal	Mediation with vs. without partner, no vs. respiration vs. EEG vs. combined feedback	EEG, Resp	Feedback, correlation stimuli comparison
Desnoyers-Stewart et al. (2019)	Virtual multiuser underwater environment with feedback about the individual and synchrony of respiration		Resp	Feedback
Lou et al. (2020)	System, which can detect facial muscle activation and transfer it to an avatar		EMG	Feedback

and defined differently. Only recently, Latoschik and Wienrich (2021) introduced a new theoretical model for VR experiences which also shows a new perspective on presence. Just as the understanding of presence evolves, so does the measurement of it.

4.1.5.2 Cybersickness

Cybersickness can be described as a set of adverse symptoms that are induced by the visual stimuli of virtual and augmented reality applications (Stauffert et al., 2020). Common symptoms include headache, dizziness, nausea, disorientation, or fatigue (Kennedy et al., 1993; LaViola, 2000). There are multiple theories on what might be the causes of cybersickness, whereas the most common revolve around sensory mismatches and postural instability (Rebenitsch and Owen, 2016).

Besides questionnaires and tests for postural instability, the assessment of the physiological state of a VR user is one of the common ways to measure cybersickness (Rebenitsch and Owen, 2016). In recent years researchers used several approaches to assess physiological measures and find out how much they correlate with cybersickness. Gavgani et al. (2017) used a virtual roller-coaster ride that subjects were asked to ride on three consecutive days. This roller-coaster ride was quite effective at inducing cybersickness as only one of fourteen subjects completed all rides while the others terminated theirs due to nausea. However, it took the participants significantly more time to abort the ride on the third day, compared to the first, which speaks for a habituation. During the 15-min rides, heart rate, respiration rate, and skin conductance were monitored and participants had to give a subjective assessment of their felt motion sickness. Results demonstrated that the nausea level of subjects continuously increased over the course of the ride. The measurement of the forehead skin conductance was the best physiological correlate to the gradually increasing nausea. A virtual roller-coaster ride was also leveraged in the study of Cebeci et al. (2019). Here, pupil dilation, heart rate, blink count, and saccades were analyzed. In this study, the average heart rate and the saccade mean speed were the highest when cybersickness symptoms occurred. Moreover, they found a correlation between the blink count, nausea and oculomotor discomfort (Kennedy et al., 1993). Approaches that use physiological data to assess cybersickness mainly use this data for the sake of comparison. This can serve to gain knowledge

TABLE 9 | Overview of the works from the field of general VR properties that were discussed in **Section 4.1.5**. The Measures column refers to the physiological measures used in the work. The entries of the Independent Variables column often do not cover everything that was considered in the work. Entries in the Purposes column refer to the categories listed in **Section 4.1.6**.

Work	Scenario	Independent variables	Measures	Purposes
Athif et al. (2020)	VR forest scenario in which the player needs to collect mushrooms that spawn randomly	Suppression of four individual presence factors vs. suppression of all factors vs. suppression of no factor	ECG, EDA EEG	Stimuli comparison
Dey et al. (2020)	Virtual cart ride through a jungle	High presence vs. low presence (manipulated through visual fidelity, embodiment, reactivity and control over the environment)	ECG, EDA EEG	Stimuli comparison
Deniaud et al. (2015)	Following another car in a virtual driving simulator	Visual realistic vs. unrealistic environment, good vs. bad visibility of the road	ECG, EDA	Stimuli comparison, correlation
Felnhofer et al. (2014)	Public speaking task in front of a virtual audience	High anxious vs. low anxious subjects	ECG	Process analysis, correlation, group comparison
Felnhofer et al. (2015)	Virtual park scenario that tries to elicit different emotions	Park that is intended to elicit joy vs. anger vs. boredom vs. sadness vs. anxiety	EDA	Stimuli comparison, correlation
da Costa et al. (2018)	Car-driving scenario with different traffic situations, tested with women with a fear of driving	Different driving sessions	ECG	Progress, process analysis, correlation
Gavgani et al. (2017)	Virtual roller-coaster ride on three consecutive days	Day of exposure, before vs. after exposure	ECG, EDA, SKT	Progress, process analysis, correlation
Cebeci et al. (2019)	Experiencing different virtual environments, i.e. campfire, hospital, and roller-coaster scene	Different scenes	PPG, ET	Stimuli comparison, correlation process analysis
Stauffert et al. (2018)	VR object search task	With vs. without induced latency jitter	PPG, EDA	Stimuli comparison
Líndal et al. (2018)	Getting from one place to another in a virtual city using different traveling methods	Driving along vs. teleportation technique	BP	Process analysis, stimuli comparison
Gersak et al. (2020)	Virtual roller-coaster ride	2D TV vs. four different VR headsets	EDA, SKT PPG, Resp	Stimuli comparison, correlation
Guna et al. (2019), Guna et al. (2020)	3D video of a beach scene and a roller-coaster ride	Neutral vs. action content, 2D TV vs. three different VR headsets vs. mobile VR	EDA, SKT PPG, Resp	Stimuli comparison, correlation
Plouzeau et al. (2018)	Navigating through a virtual forest, whereat acceleration parameters can be adjusted to the EDA of a subject	With vs. without adaption of acceleration	EDA	Adaption, stimuli comparison

about the connection of unpleasant VR experiences and latency jitter (Stauffert et al., 2018), navigation techniques (Líndal et al., 2018), or the display type (Guna et al., 2019; Gersak et al., 2020; Guna et al., 2020). Plouzeau et al. (2018) used cybersickness indicators in an adaption mechanism for their VR application. They introduced a navigation method that allows the user to move and rotate in the virtual environment with the help of two joysticks. The acceleration of the navigation is adapted according to an objective indicator for simulator sickness, i.e. EDA. When the EDA increases the acceleration decreases proportionally and vice versa.

A summary of the works discussed in this section can be found in **Table 9**.

4.1.6 High-Level Purposes

Throughout this section, we gave an overview of the usage of physiological measures in VR to assess the state of the user. We listed fields of application and concretely explained how physiological measures are used in them. Across the fields of application, physiological measures are used for recurring purposes. To summarize this overview we turn to the meta-level to highlight these recurring themes for the usage of physiological data in VR. The categories are not mutually exclusive and are not always clearly separable.

- Stimuli Comparison: Physiological measures can be used to determine how the response to a virtual stimulus compares to the response to another (virtual) stimulus. In these cases the independent variable is the stimulus and the dependent variable is the physiological measure. Examples include works that compare responses to real life situations and their virtual counterparts (Chang et al., 2019; Syrjamaki et al., 2020). Others compare how different kinds of virtual audiences impact stress responses (Barreda-Angeles et al., 2020; Mostajeran et al., 2020).
- Group Comparison: Physiological measures can be used to determine how the response to the same stimulus compares between groups of people. In these studies the independent variable is the user group and the dependent variable is the physiological measure. Examples include works that compare phobic with non-phobic subjects (Breuninger et al., 2017; Kishimoto and Ding, 2019) or subjects with and without autism (Simões et al., 2018).
- Process analysis: Physiological measures can be used to determine how the response changes over the course of a virtual simulation. In these cases the independent variable is the time of measurement and the dependent variable is the physiological measure. Thus, the effect of the appearance of a certain stimulus can be determined, e.g. a knife attack

(González-Franco et al., 2014) or a noise burst (Mueller et al., 2017).

- Progress: Physiological measures can be used to determine a change in response to the same stimulus throughout multiple expositions. In these studies the independent variable is the number of expositions or sessions and the dependent variable is the physiological measure. This is often done to quantify the progress of a therapy or training (Lee et al., 2015; Shiban et al., 2017) but can, for example, also be used to determine a habituation to cybersickness inducing stimuli (Gavgani et al., 2017).
- Correlation: Physiological measures can be used to establish a relationship between the measure and a second variable. Usually, both measures are dependent variables of the research setup. Typical examples assess the relationship between physiological and subjective measurements, e.g. of embodiment (Alchalabi et al., 2019) or cybersickness (John, 2019).
- Classification: Physiological measures can be used to differentiate users based on the response to a virtual stimulus. The goal of these approaches is to determine if the information in the physiological data is sufficient to reflect the changes in the independent variable. Examples can be the classification of specific groups of people, e.g. healthy and addicted people (Ding X. et al., 2020) or people under low and high stress (Ishaque et al., 2020).
- Feedback: Physiological measures can be presented to the user or a second person to make latent and unconscious processes visible. This is particularly common in relaxation applications where the stress level can be visualized for the user (Patibanda et al., 2017; Blum et al., 2019) but it can also be used to inform the supervisor of a therapy or training session about the user's condition (Bayan et al., 2018; Streck et al., 2019). This purpose differs from the previous ones in that the physiological measurement is no longer intended to indicate the manipulation of an independent variable.
- Adaption: Physiological measurements can be used to adapt the system status to the state of the user. A typical example is the adaption of training and therapy systems based on effort and stress indicators (Campbell and Fraser, 2019; Bălan et al., 2020). This is similar to the feedback purpose in that the measurements here are used to make changes to the system and not to allow comparisons. While feedback approaches are really just focused on visualizing the physiological data, here it is more about changing the behavior of the application.

4.2 Part 2: Characteristics of Experience and Their Measurement in Virtual Reality

In the second part of the discussion, we focus on the results of the search for classification approaches depicted in **Table 2**. Here, we discuss approaches that expose participants to a particular VR stimulus that is known to trigger a particular characteristic of experience. The focus of the studies is on how well this manipulation is reflected by the physiological measurements. We use those classification approaches to show which measures, sensors, and algorithms have been used to gauge the targeted characteristic of experience. A universal solution to measure and interpret those specific experiences does not exist, as this is usually context dependent. So what this work cannot do is to give strict guidelines for which measures should be used for which case. The field is too diverse and the focus of the work too broad.

A comparison of the accuracy of the specific approaches, should be treated with caution as they are partly obtained under different circumstances. Results show more of a rough guide to how well the classification works and should not be compared 1-to-1. All the classifiers reported here are, in principle, successful in distinguishing different levels of an experience. This means all examples show combinations of signals, sensors, and algorithms that can work for the assessment of experience in VR.

Our review of classification approaches showed that in immersive VR there are some main characteristics of experience that are predominantly assessed with the help of physiological data. These experiences are arousal, valence, stress, anxiety, and cognitive workload. Those constructs are similar and interrelated. Stress and anxiety can be seen as a form of hyperarousal and cognitive workload itself can be seen as a stress factor (Gaillard, 1993; Blanco et al., 2019). Nevertheless, most of the works focus on one of the characteristics of experience and they have different approaches to elicit and assess them. The discussion is separated according to these characteristics of experience. The reader should still keep in mind that the constructs are related.

4.2.1 Arousal and Valence

Studies from this domain usually base their work on the Circumplex Model of Affects (Russell and Mehrabian, 1977; Posner et al., 2005). This model arranges human emotions in a two-dimensional coordinate system. One axis of this coordinate system represents arousal, i.e. the activation of the neural system, and one axis represents valence, i.e. how positive or negative an emotion is perceived. Hence, classifiers from this category usually distinguished high and low levels of arousal or positive and negative valence. Arousal inducing scenes often comprise a virtual roller-coaster ride (Hofmann et al., 2018; Teo and Chia, 2018; Bilgin et al., 2019) or dynamic mini-games (Shumailov and Gunes, 2017; Ding Y. et al., 2020). Emotional scenes are often used to manipulate the valence of people (Shumailov and Gunes, 2017; Mavridou et al., 2018b; Zheng et al., 2020). Such scenes can be taken from a database (Samson et al., 2016) or be tested in a pre-study to see what emotions they trigger (Zhang W. et al., 2017).

The most commonly used physiological measure for the classification of arousal and valence is EEG. The trend here seems to be towards the more comfortable wearable EEG sensors, e.g. a EEG headset (Teo and Chia, 2018; Bilgin et al., 2019; Ding Y. et al., 2020; Suhaimi et al., 2020) or textile electrodes inside the HMD (Xu et al., 2019). Some works use cardiovascular data next to the EEG information (Marín-Morales et al., 2018; Mavridou et al., 2018b). Deviating from the EEG approach, Zheng et al. (2020) leveraged pupillometry and Shumailov and Gunes (2017) forearm EMG to classify arousal and valence. Both examples also worked with comfortable and easy-to-setup sensors.

The deep neural network for the two-level classification of emotional arousal of Teo and Chia (2018) achieved an accuracy of 96.32% in a 10-fold cross validation. This result was achieved, just with the data from the Muse 4-channel EEG headband. For a binary valence classification Shumailov and Gunes (2017) reported an F1 value of 0.85. This value was achieved with the help of a support vector machine and EMG armband data, captured while playing VR games. The highest value for classifying arousal and valence at the same time (four classes) comes from Suhaimi et al. (2020). Their random-forest classifier achieved an accuracy of 82.49% in a 10-fold cross validation, distinguishing four different emotions that are embedded in the valence-arousal model.

When a researcher wants to assess arousal in a virtual environment EEG signals appear to be the go-to indicators. In addition, cardiovascular data also appears to be useful for this purpose. Six out of the eight presented arousal classifiers successfully used one or both of the signals to distinguish high and low arousal in the virtual realm. The systematic review of Marín-Morales et al. (2020) about the recognition of emotions in VR generally confirms this impression. They list sixteen works that assessed arousal in VR, whereat fifteen of them used EEG or heart rate variability signals. However, the review of Marín-Morales et al. (2020) also shows that nine of the sixteen works used EDA data to estimate arousal, a signal that did not appear among recent classification algorithms. One reason for this could be that, among the works listed by Marín-Morales et al. (2020), the older ones tended to leverage EDA for arousal assessment, and this work here just considers literature from the last few years. Nevertheless, this does not mean that EDA measurements are not important for estimating arousal anymore. Only recently, Granato et al. (2020) found the skin conductance level to be one of the most informative features when it comes to the assessment of arousal. Also worth mentioning is the work of Shumailov and Gunes (2017) which showed that also forearm EMG is suitable for the classification of arousal levels. They showed this in a setup where subjects moved a lot as they were playing VR games, while other approaches usually gather their data in a setup where subjects must remain still. Due to movement artifacts, it is questionable to what extent the other classifiers are transferable to setups that include a lot of motion. As for the sensors, various works showed that EEG data collected with easy-to-use headsets is sufficient to distinguish arousal levels in VR (Teo and Chia, 2018; Bilgin et al., 2019; Ding Y. et al., 2020).

The differentiation of positive and negative valence appears to be similar to arousal. Our results showed that most frequently EEG data was used for its assessment. Also notable is the attempt to classify arousal based on facial expressions. Even if an HMD is worn, this is possible through facial EMG (Mavridou et al., 2018a).

4.2.2 Stress

Studies that work on the classification of stress often used some kind of dynamic or unpredictable virtual environment to elicit the desired responses, e.g. a roller-coaster ride (Ishaque et al., 2020) or a guard, patrolling in a dark room (Ham et al., 2017). Stress is usually regulated with an additional assignment, e.g. an arithmetic task (Cho et al., 2017) or a Stroop task (Ishaque et al., 2020).

Looking at the signals with which stress was attempted to be classified, it is noticeable that each approach measures the cardiovascular activity. Either with optical sensors on the finger (Cho et al., 2017; Ham et al., 2017) or with electrical sensors (Tartarisco et al., 2015; Robitaille and McGuffin, 2019; Ishaque et al., 2020). Additional measures that were used by these studies are EDA (Cho et al., 2017; Ishaque et al., 2020), skin temperature (Cho et al., 2017), respiration Ishaque et al. (2020), or motion activity (Tartarisco et al., 2015; Robitaille and McGuffin, 2019).

The kernel-based extreme learning machine of Cho et al. (2017) distinguishes five stress levels and it achieved an accuracy of over 95% in a leave-one-out cross validation. Their classifier was trained with PPG, EDA, and skin temperature signals that were gathered relatively simple with four finger electrodes. In an even simpler setup, with only one finger-worn PPG sensor and a Linear Discriminant Analysis, Ham et al. (2017) achieved an accuracy of approximately 80% for three different classes. Tartarisco et al. (2015) took an approach with a wearable chest band. They collected ECG, respiration, and motion data and trained a neuro-fuzzy neural network that achieved an accuracy of 83% for four different classes.

Traditionally, heart rate variability is regarded as one of the most important indicators of stress (Melillo et al., 2011; Kim et al., 2018). This coincides with our results as the most commonly used signals for stress classification were PPG and ECG. This impression is also confirmed when considering non-classification approaches in VR. For example, if one looks at the VR adaptations of the Trier Social Stress Test mentioned in **Section 4.1.4**, one finds that in all the listed examples the heart activity is measured. Two other signals frequently used in research to indicate changes in stress level is EDA (Kurniawan et al., 2013; Anusha et al., 2017; Bhoja et al., 2020) and skin temperature (Vinkers et al., 2013; Herborn et al., 2015). Both signals and heart rate variability were compared by Cho et al. (2017) in VR. Results indicated that PPG and EDA provided more information about the stress level of the immersed people than the skin temperature, whereas PPG features were best suited for the distinction of stress. The combination of EDA and cardiovascular data seems to be a good compromise for measuring stress in VR.

Our results also suggest that the better classification results were achieved with the help of more obtrusive sensors like multiple electrodes on the fingers or the body (Cho et al., 2017; Ishaque et al., 2020). This is somewhat problematic as a lot of VR scenarios require quite some movement interaction. Individual electrodes distributed over the body could be bothersome. Approaches with more comfortable chest bands showed somewhat worse accuracy values, yet were able to effectively classify different levels of stress (Tartarisco et al., 2015; Robitaille and McGuffin, 2019). Future research could aim on improving the quality of stress indicators in VR based on unobtrusive sensors. In addition to chest bands, wrist-worn devices could be used given that they can deliver the important cardiovascular and EDA signals. Indeed, the focus in these scenarios is only on creating stress. Relaxation environments that do the opposite could also provide data to train future classifiers.

4.2.3 Cognitive Workload

As the name suggests, VR studies that work on the estimation of cognitive workload often used mentally demanding tasks that allow for a manipulation with different levels of difficulty. This can be abstract assignments like the n-back task (Tremmel et al., 2019; Tremmel, 2020) or a cube puzzle (Collins et al., 2019), but also more concrete scenarios like a flight simulator with different difficulties (Kakkos et al., 2019). It is in these scenarios that cognitive workload differs from the other characteristics of experiences listed here. While the other experiences can be placed somewhere in the Circumplex Model of Affects and therefore have an emotional character, the focus here is on a mental effort that must be performed by the subjects. In contrast to the stress simulations, here it is purely a matter of the cognitive demands of the tasks and not on environmental factors that are supposed to create additional stress.

Most frequently cognitive workload classifiers worked with an EEG signal (Kakkos et al., 2019; Siravenha et al., 2019; Tremmel et al., 2019; Tremmel, 2020). An exception to this is the work of (Collins et al., 2019) who approached the classification of workload in VR with PPG and EDA signals.

It is also (Collins et al., 2019) who reached the highest accuracy among the cognitive workload classifiers that are listed here. Based on information about the cardiovascular activity, collected with a wristband, they created a random forest classifier that predicts three different levels of cognitive workload with an accuracy of 91.75%. Among the EEG based approaches, Kakkos et al. (2019) report the highest accuracy. With data from 64-scalp electrodes, they trained a linear discriminant analysis classifier that reached an accuracy of 89% for a prediction of three different workload levels.

Older studies established heart rate features as the most reliable predictors of cognitive workload (Hancock et al., 1985; Vogt et al., 2006). More recent works argue for EEG data as the most promising signal for classifying workload (Christensen et al., 2012; Hogervorst et al., 2014). This trend is also visible in our results, as almost all of the classifiers for workload used EEG. However, Collins et al. (2019) showed that a cognitive workload classification in VR can also work with PPG signals. So both, heart rate and EEG features seem to be usable for workload classification in VR. A recent review on the usage of physiological data to assess cognitive workload also shows that cardiovascular and EEG data are two main measures for this purpose (Charles and Nixon, 2019). Charles and Nixon (2019) report that the second most used signal is the assessment of cognitive workload are ocular measures, i.e. blink rate and pupil size. Those measures did not appear at all in the classifiers for workload that we found. Closing this gap could be a task for future research, especially because of the availability of sensors that allow capturing pupillometry data inside an HMD.

Regarding the sensors, we found that all the EEG devices that were used for a workload classification were quite cumbersome (caps with multiple wet electrodes). The classification with more comfortable devices like the Emotiv Epoc or the Muse headset is still pending. When using pulse sensors, Collins et al. (2019) already showed that the data from a convenient wristband can be sufficient to distinguish workload levels, however, more examples are needed to confirm this impression.

4.2.4 Anxiety

The classification of anxiety is closely related to the virtual exposure therapies presented in **Section 4.1.1**. This becomes particularly clear when one considers the scenarios in which the data for the classifiers were gathered. The scenario is either the exposure to different altitudes (Hu et al., 2018; Wang et al., 2018; Bălan et al., 2020) or a speech in front of a crowd (Salkevicius et al., 2019).

The studies of Hu et al. (2018) and Wang et al. (2018) work with more cumbersome sensors, i.e. over 30 scalp electrodes, for capturing EEG data. The convolutional neural network of Hu et al. (2018) reached an accuracy of 88.77% in a 10-fold cross validation when classifying four different levels of acrophobia. The support vector machine of Wang et al. (2018) reached an accuracy of 96.20% in a 5-fold cross validation, yet only distinguished three levels of fear.

Salkevicius et al. (2019) present a VR anxiety classification based on a wearable sensor. With the help of the Empatica E4 wristband sensor, they collected PPG, EDA, and skin temperature data. They created a fusion-based support vector machine that classifies four different levels of anxiety. In a 10× 10-fold cross validation it reached an accuracy of 86.10%, which is comparable to what Hu et al. (2018) achieved with a more elaborate 30 electrodes setup.

Anxiety is usually characterized by sympathetic activation. Therefore, in the past many studies have found correlations between anxiety levels and numerous features of cardiovascular activity and EDA measurements (Kreibig, 2010). In VR applications, too, most researchers use heart rate variability and EDA data to make anxiety measurable (Marín-Morales et al., 2020). In our results, however, this combination only appeared in the study of Salkevicius et al. (2019). From this work, it can be concluded that heart rate variability, EDA, and skin temperature data are in general suitable for distinguishing different anxiety levels in VR. Moreover, it showed that the fusion of these three signals can considerably increase the quality of the prediction, which is particularly useful when using a wristband that can conveniently deliver this data like the Empatica E4.

Our results indicate the suitability of EEG data as a sensitive measure for anxiety in VR. Each of the fear-related approaches, except that of Salkevicius et al. (2019), used knowledge of the brain activity for classification. Additionally, the combination with cardiovascular measures seems to work fine (Balan et al., 2019; Bălan et al., 2020). As with cognitive workload, the EEG data here has been mainly captured with comprehensive electrode setups. Future work could seek for classification with the more comfortable headsets. Future anxiety classification approaches could also include EMG signals from the orbicularis oculi muscle. This measure can serve to identify startle responses (Maples-Keller et al., 2019; Mertens et al., 2019).

4.2.5 Other Classifiers

We also found classification approaches for more seldom assessed characteristics of experience. Orlosky et al. (2019) built a classifier that predicts if a learner in a virtual environment understood a given term or not. Based on data of eye movement and pupil size, they report a classification accuracy of 75%. Understanding is also a focus in the study of Collins et al. (2019). They use EDA information to recognize a moment of insight (Aha! moment). Also, the severity of cybersickness can be classified with physiological data. Jeong et al. (2019) used an Emotiv Epoc+ EEG headset to capture data for implementing a neural network. This network was able to detect if someone feels sick or not with an accuracy of 98.82%. Just like cybersickness, the visual fatigue caused by an HMD is quite specific to VR. Wang Y. et al. (2019) built a classifier that could distinguish two levels of visual fatigue with an accuracy of 90.79%.

4.3 Limitations

Although this review provides a fairly comprehensive overview of the usage of physiological signals in VR, it is not without limitations. Of course, there are a variety of application areas for physiological data in VR that we have not addressed. Indeed, we have only reported a fraction of the papers that were left after the screening process. The scope of this review limits us to only a superficial discussion of the specific field. To generate a deeper understanding one would have to dedicate a separate review to many of the topics. Moreover, we only focused on works that used physiological measures to gauge the state of the user. However, the measurements can also be used to make active system commands, for example with the help of a brain-computer interface. Additionally, our discussion of classification approaches could only cover certain areas. We discussed the characteristics of experience and the signals with which they were assessed. We have not discussed the features of the specific signals.

5 CONCLUSION

The use of physiological measures in VR is very wide and versatile. In the first part of this review, we provided a structured overview of the field. We showed how physiological signals are used in therapy, training and entertainment applications as well as the usage with functional and general VR properties. We also highlighted how the knowledge obtained through physiological data is used. This ranges from the comparison of different stimuli over the adaptation of the virtual environment to statistical methods such as correlation. In the second part, we focused on classification approaches that can show which characteristics of experience can be assessed with which measures and sensors. Approaches for the classification of arousal, valence, anxiety, stress, and cognitive workload were most prominent. EEG and cardiovascular data were most commonly used for the assessment of those dimensions. In many areas, simple and easy-to-use sensors were sufficient to distinguish different levels of an experience.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

AH conducted the search, processing, and ordering of the literature and took the lead in writing. ML has contributed to the categorization of the literature and the overall structuring of this review. He is also the supervisor of this project.

FUNDING

This research has been funded by the German Federal Ministry of Education and Research in the project VIA-VR (project number 16SV8444).

ACKNOWLEDGMENTS

Special thanks to Sebastian Oberdörfer and Samantha Monty for proofreading and feedback and to Murat Yalcin for assisting in machine learning questions.

REFERENCES

Ahmaniemi, T., Lindholm, H., Muller, K., and Taipalus, T. (2017). "Virtual Reality Experience as a Stress Recovery Solution in Workplace," in 2017 IEEE Life Sciences Conference (LSC), 206–209.

Al-Khalidi, F. Q., Saatchi, R., Burke, D., Elphick, H., and Tan, S. (2011). Respiration Rate Monitoring Methods: A Review. *Pediatr. Pulmonol.* 46, 523–529. doi:10.1002/ppul.21416

Alchalabi, B., Faubert, J., and Labbe, D. R. (2019). "EEG Can Be Used to Measure Embodiment when Controlling a Walking Self-Avatar," in 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), 776–783.

Alian, A. A., and Shelley, K. H. (2014). Photoplethysmography. *Best Pract. Res. Clin. Anaesthesiol.* 28, 395–406. doi:10.1016/j.bpa.2014.08.006

Almeida, J., Spofford, C., van 't Wout, M., Unger, W., Philip, N., Carpenter, L., et al. (2016). Heart Rate Variability Responses to a Standardized Virtual Reality Exposure in Veterans With PTSD. *Neuropsychopharmacology* 41, S460–S461. doi:10.1038/npp.2016.242

Anderson, A. P., Mayer, M. D., Fellows, A. M., Cowan, D. R., Hegel, M. T., and Buckey, J. C. (2017). Relaxation with Immersive Natural Scenes Presented Using Virtual Reality. *Aerosp. Med. Hum. Perform.* 88, 520–526. doi:10.3357/AMHP.4747.2017

Anusha, A., Joy, J., Preejith, S., Joseph, J., and Sivaprakasam, M. (2017). "Differential Effects of Physical and Psychological Stressors on Electrodermal Activity," in 2017

39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (IEEE), 4549–4552.

Armel, K. C., and Ramachandran, V. S. (2003). Projecting Sensations to External Objects: Evidence from Skin Conductance Response. *Proc. R. Soc. Lond. B* 270, 1499–1506. doi:10.1098/rspb.2003.2364

Athif, M., Rathnayake, B. L. K., Nagahapitiya, S. M. D. B. S., Samarasinghe, S. A. D. A. K., Samaratunga, P. S., Peiris, R. L., et al. (2020). "Using Biosignals for Objective Measurement of Presence in Virtual Reality Environments," in 2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC), 3035–3039.

Balan, O., Moise, G., Moldoveanu, A., Moldoveanu, F., and Leordeanu, M. (2019). "Automatic Adaptation of Exposure Intensity in vr Acrophobia Therapy, Based on Deep Neural Networks," in 27th European Conference on Information Systems - Information Systems for a Sharing Society, ECIS 2019, Stockholm and Uppsala, Sweden, June 8–14, 2019.

Barreda-Ángeles, M., Aleix-Guillaume, S., and Pereda-Baños, A. (2020). Users' Psychophysiological, Vocal, and Self-Reported Responses to the Apparent Attitude of a Virtual Audience in Stereoscopic 360°-Video. *Virtual Real.* 24, 289–302. doi:10.1007/s10055-019-00400-1

Barsade, S. G., Ramarajan, L., and Westen, D. (2009). Implicit Affect in Organizations. *Res. Organ. Behav.* 29, 135–162. doi:10.1016/j.riob.2009.06.008

Bayan, S., Assaf, K., Yassin, M., Cherry, A., Raad, M., and Hamawy, L. (2018). "A Virtual Reality Assisted Rehabilitation System for Physical Therapy," in 2018 International Conference on Computer and Applications (ICCA), 32–40.

Bekele, E., Wade, J., Bian, D., Fan, J., Swanson, A., Warren, Z., et al. (2016). "Multimodal Adaptive Social Interaction in Virtual Environment (masi-vr) for Children with Autism Spectrum Disorders (Asd)," in 2016 IEEE Virtual Reality (VR) (IEEE), 121–130.

Benedek, M., and Kaernbach, C. (2010). A Continuous Measure of Phasic Electrodermal Activity. *J. Neurosci. Methods* 190, 80–91. doi:10.1016/j.jneumeth.2010.04.028

Betka, S., Canzoneri, E., Adler, D., Herbelin, B., Bello-Ruiz, J., Kannape, O. A., et al. (2020). Mechanisms of the Breathing Contribution to Bodily Self-Consciousness in Healthy Humans: Lessons from Machine-Assisted Breathing?. *Psychophysiology* 57, e13564. doi:10.1111/psyp.13564

Bhoja, R., Guttman, O. T., Fox, A. A., Melikman, E., Kosemund, M., and Gingrich, K. J. (2020). Psychophysiological Stress Indicators of Heart Rate Variability and Electrodermal Activity With Application in Healthcare Simulation Research. *Sim. Healthc.* 15, 39–45. doi:10.1097/sih.0000000000000402

Bian, Y., Yang, C., Gao, F., Li, H., Zhou, S., Li, H., et al. (2016). A Framework for Physiological Indicators of Flow in VR Games: Construction and Preliminary Evaluation. *Pers. Ubiquitous Comput.* 20, 821–832. doi:10.1007/s00779-016-0953-5

Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K.-M., and Robbins, K. A. (2015). The Prep Pipeline: Standardized Preprocessing for Large-Scale Eeg Analysis. *Front. Neuroinform.* 9, 16. doi:10.3389/fninf.2015.00016

Bilgin, P., Agres, K., Robinson, N., Wai, A. A. P., and Guan, C. (2019). "A Comparative Study of Mental States in 2D and 3D Virtual Environments Using EEG," in 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), 2833–2838.

Blanco, J. A., Vanleer, A. C., Calibo, T. K., and Firebaugh, S. L. (2019). Single-Trial Cognitive Stress Classification Using Portable Wireless Electroencephalography. *Sensors* 19, 499. doi:10.3390/s19030499

Blum, J., Rockstroh, C., and Goeritz, A. S. (2019). Heart Rate Variability Biofeedback Based on Slow-Paced Breathing With Immersive Virtual Reality Nature Scenery. *Front. Psychol.* 10, 2172. doi:10.3389/fpsyg.2019.02172

Botvinick, M., and Cohen, J. (1998). Rubber Hands 'Feel' Touch that Eyes See. *Nature* 391, 756. doi:10.1038/35784

Bozkir, E., Geisler, D., and Kasneci, E. (2019). "Person Independent, Privacy Preserving, and Real Time Assessment of Cognitive Load Using Eye Tracking in a Virtual Reality Setup," in 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), 1834–1837.

Bradley, M. M., and Lang, P. J. (1994). Measuring Emotion: the Self-Assessment Manikin and the Semantic Differential. *J. Behav. Ther. Exp. Psychiatry* 25, 49–59. doi:10.1016/0005-7916(94)90063-9

Braithwaite, J. J., Watson, D. G., Jones, R., and Rowe, M. (2013). A Guide for Analysing Electrodermal Activity (eda) & Skin Conductance Responses (scrs) for Psychological Experiments. *Psychophysiology* 49, 1017–1034.

Breuninger, C., Slama, D. M., Kraemer, M., Schmitz, J., and Tuschen-Caffier, B. (2017). Psychophysiological Reactivity, Interoception and Emotion Regulation in Patients With Agoraphobia during Virtual Reality Anxiety Induction. *Cogn. Ther. Res.* 41, 193–205. doi:10.1007/s10608-016-9814-9

Browning, M. H. E. M., Mimnaugh, K. J., van Riper, C. J., Laurent, H. K., and LaValle, S. M. (2019). Can Simulated Nature Support Mental Health? Comparing Short, Single-Doses of 360-Degree Nature Videos in Virtual Reality With the Outdoors. *Front. Psychol.* 10, 2667. doi:10.3389/fpsyg.2019.02667

Bruder, L., and Peters, J. (2020). A Virtual Reality Setup for the Combined Assessment of Subjective, Behavioral and Psychophysiological Cue-Reactivity in Gambling Disorder. *bioRxiv*. doi:10.1101/2020.08.07.237826

Bălan, O., Moise, G., Moldoveanu, A., Leordeanu, M., and Moldoveanu, F. (2020). An Investigation of Various Machine and Deep Learning Techniques Applied in Automatic Fear Level Detection and Acrophobia Virtual Therapy. *Sensors* 20, 496. doi:10.3390/s20020496

Cairns, P. E., and Cox, A. L. (2008). *Research Methods for Human-Computer Interaction*. Cambridge, UK: Cambridge University Press.

Calabro, R. S., Naro, A., Russo, M., Leo, A., De Luca, R., Balletta, T., et al. (2017). The Role of Virtual Reality in Improving Motor Performance as Revealed by EEG: A Randomized Clinical Trial. *J Neuroeng. Rehabil.* 14, 53. doi:10.1186/s12984-017-0268-4

Campbell, J., and Fraser, M. (2019). "Switching it up: Designing Adaptive Interfaces for Virtual Reality Exergames," in Proceedings of the 31st European Conference on Cognitive Ergonomics (New York, NY: Association for Computing Machinery), 177–184.

Carreiras, C., Alves, A. P., Lourenço, A., Canento, F., Silva, H., Fred, A., et al. (2015). BioSPPy: Biosignal Processing in Python. Dataset.

Castelvecchi, D. (2016). Low-Cost Headsets Boost Virtual Reality's Lab Appeal. *Nature* 533, 153–154. doi:10.1038/533153a

Cebeci, B., Celikcan, U., and Capin, T. K. (2019). A Comprehensive Study of the Affective and Physiological Responses Induced by Dynamic Virtual Reality Environments. *Comput. Animat. Virtual Worlds* 30, e1893. doi:10.1002/cav.1893

Chang, T. P., Beshay, Y., Hollinger, T., and Sherman, J. M. (2019). Comparisons of Stress Physiology of Providers in Real-Life Resuscitations and Virtual Reality-Simulated Resuscitations. *Simul. Healthc.* 14, 104–112. doi:10.1097/SIH.0000000000000356

Charles, R. L., and Nixon, J. (2019). Measuring Mental Workload Using Physiological Measures: A Systematic Review. *Appl. Ergon.* 74, 221–232. doi:10.1016/j.apergo.2018.08.028

Cho, D., Ham, J., Oh, J., Park, J., Kim, S., Lee, N.-K., et al. (2017). Detection of Stress Levels from Biosignals Measured in Virtual Reality Environments Using a Kernel-Based Extreme Learning Machine. *Sensors* 17, 2435. doi:10.3390/s17102435

Christensen, J. C., Estepp, J. R., Wilson, G. F., and Russell, C. A. (2012). The Effects of Day-to-Day Variability of Physiological Data on Operator Functional State Classification. *NeuroImage* 59, 57–63. doi:10.1016/j.neuroimage.2011.07.091

Clifford, R. M. S., Engelbrecht, H., Jung, S., Oliver, H., Billinghurst, M., Lindeman, R. W., et al. (2020). Aerial Firefighter Radio Communication Performance in a Virtual Training System: Radio Communication Disruptions Simulated in VR for Air Attack Supervision. *Vis. Comp.*. doi:10.1007/s00371-020-01816-6

Clifford, R. M. S., Hoermann, S., Marcadet, N., Oliver, H., Billinghurst, M., and Lindeman, R. W. (2018). "Evaluating the Effects of Realistic Communication Disruptions in VR Training for Aerial Firefighting," in 2018 10th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games), 1–8.

Collins, J., Regenbrecht, H., Langlotz, T., Can, Y. S., Ersoy, C., and Butson, R. (2019). "Measuring Cognitive Load and Insight: A Methodology Exemplified in a Virtual Reality Learning Context," in 2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Beijing, China, 351–362.

Corbetta, P. (2003). *Social Research: Theory, Methods and Techniques*. Sage.

Currie, J., Bond, R. R., McCullagh, P., Black, P., Finlay, D. D., Gallagher, S., et al. (2019). Wearable Technology-Based Metrics for Predicting Operator Performance during Cardiac Catheterisation. *Int. J. Comput. Assist. Radiol. Surg.* 14, 645–657. doi:10.1007/s11548-019-01918-0

Czub, M., and Kowal, M. (2019). Respiration Entrainment in Virtual Reality by Using a Breathing Avatar. *Cyberpsychol. Behav. Soc. Netw.* 22, 494–499. doi:10.1089/cyber.2018.0700 da Costa, R. T., de Carvalho, M. R., Ribeiro, P., and Nardi, A. E. (2018). Virtual Reality Exposure Therapy for Fear of Driving: Analysis of Clinical Characteristics, Physiological Response, and Sense of Presence. *Braz. J. Psychiatry* 40, 192–199. doi:10.1590/1516-4446-2017-2270

Dash, A., Yadav, A., and Lahiri, U. (2019). "Physiology-Sensitive Virtual Reality Based Strength Training Platform for Post-Stroke Grip Task," in 2019 IEEE EMBS International Conference on Biomedical Health Informatics (BHI), 1–4.

De Asis, K. M. R., Guillem, E. J. P., Reyes, F. A. M., and Samonte, M. J. C. (2020). Serenity: A Stress-Relieving Virtual Reality Application Based on Philippine Environmental Variables. In Proceedings of the 2020 The 6th International Conference on Frontiers of Educational Technologies (New York, NY: Association for Computing Machinery), 155–159.

De Luca, C. (2006). *Electromyography. Encyclopedia of Medical Devices and Instrumentation*, Hoboken, USA: Wiley Online Library.

Debska, M., Polechonski, J., Mynarski, A., and Polechonski, P. (2019). Enjoyment and Intensity of Physical Activity in Immersive Virtual Reality Performed on Innovative Training Devices in Compliance with Recommendations for Health. *Int. J. Environ. Res. Public Health* 16, 3673. doi:10.3390/ijerph16193673

Delahaye, M., Lemoine, P., Cartwright, S., Deuring, G., Beck, J., Pflueger, M., et al. (2015). Learning Aptitude, Spatial Orientation and Cognitive Flexibility Tested in a Virtual Labyrinth after Virtual Stress Induction. *BMC Psychol.* 3, 22. doi:10.1186/s40359-015-0080-5

Delvigne, V., Ris, L., Dutoit, T., Wannous, H., and Vandeborre, J.-P. (2020). "Vera: Virtual Environments Recording Attention," in 2020 IEEE 8th International Conference on Serious Games and Applications for Health (SeGAH) (IEEE), 1–7.

Deniaud, C., Honnet, V., Jeanne, B., and Mestre, D. (2015). "An Investigation into Physiological Responses in Driving Simulators: An Objective Measurement of Presence," in 2015 Science and Information Conference (SAI), 739–748.

Desnoyers-Stewart, J., Stepanova, E. R., Pasquier, P., and Riecke, B. E. (2019). "JeL: Connecting through Breath in Virtual Reality," in Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems CHIEA'19, Glasgow, United Kingdom (New York, NY: Association for Computing Machinery), 1–6.

Detez, L., Greenwood, L.-M., Rave, R. S., Wilson, E., Chandler, T., Ries, T., et al. (2019). A Psychophysiological and Behavioural Study of Slot Machine Near-Misses Using Immersive Virtual Reality. *J. Gambl. Stud.* 35, 929–944. doi:10.1007/s10899-018-09822-z

Dey, A., Chatburn, A., and Billinghurst, M. (2019a). "Exploration of an EEG-Based Cognitively Adaptive Training System in Virtual Reality," in 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), 220–226.

Dey, A., Chen, H., Hayati, A., Billinghurst, M., and Lindeman, R. W. (2019b). "Sharing Manipulated Heart Rate Feedback in Collaborative Virtual Environments," in 2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), 248–257.

Dey, A., Chen, H., Zhuang, C., Billinghurst, M., and Lindeman, R. W. (2018). "Effects of Sharing Real-Time Multi-Sensory Heart Rate Feedback in Different Immersive Collaborative Virtual Environments," in 2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), 165–173.

Dey, A., Phoon, J., Saha, S., Dobbins, C., and Billinghurst, M. (2020). "Neurophysiological Effects of Presence in Calm Virtual Environments," in 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), 744–745.

Ding, J., He, Y., Chen, L., Zhu, B., Cai, Q., Chen, K., et al. (2019). Virtual Reality Distraction Decreases Pain during Daily Dressing Changes Following Haemorrhoid Surgery. *J. Int. Med. Res.* 47, 4380–4388. doi:10.1177/0300060519857862

Ding, N., Zhou, W., and Fung, A. Y. H. (2018). Emotional Effect of Cinematic VR Compared with Traditional 2D Film. *Telemat. Inform.* 35, 1572–1579. doi:10.1016/j.tele.2018.04.003

Ding, X., Li, Y., Li, D., Li, L., and Liu, X. (2020). Using Machine-Learning Approach to Distinguish Patients with Methamphetamine Dependence from Healthy Subjects in a Virtual Reality Environment. *Brain Behav.* 10, e01814. doi:10.1002/brb3.1814

Ding, Y., Robinson, N., Zeng, Q., Chen, D., Wai, A. A. P., Lee, T.-S., et al. (2020). "TSception:A Deep Learning Framework for Emotion Detection Using EEG," in 2020 International Joint Conference on Neural Networks (IJCNN), 1–7.

Döllinger, N., Wienrich, C., and Latoschik, M. E. (2021). Challenges and Opportunities of Immersive Technologies for Mindfulness Meditation: A Systematic Review. *Front. Virt. Real.* 2, 29. doi:10.3389/frvir.2021.644683

Drolet, M., Yumbla, E. Q., Hobbs, B., and Artemiadis, P. (2020). "On the Effects of Visual Anticipation of Floor Compliance Changes on Human Gait: Towards Model-Based Robot-Assisted Rehabilitation," in 2020 IEEE International Conference on Robotics and Automation (ICRA), 9072–9078.

Dunning, D., Heath, C., and Suls, J. M. (2004). Flawed Self-Assessment: Implications for Health, Education, and the Workplace. *Psychol. Sci. Public Interest* 5, 69–106. doi:10.1111/j.1529-1006.2004.00018.x

Ehgoetz Martens, K. A., Ellard, C. G., and Almeida, Q. J. (2015). Anxiety-Provoked Gait Changes are Selectively Dopa-Responsive in Parkinson's Disease. *Eur. J. Neurosci.* 42, 2028–2035. doi:10.1111/ejn.12928

Ehgoetz Martens, K. A., Ellard, C. G., and Almeida, Q. J. (2016). Evaluating the Link Between Dopaminergic Treatment, Gait Impairment, and Anxiety in Parkinson's Disease. *Move. Disord. Clin. Pract.* 3, 389–394. doi:10.1002/mdc3.12298

Ehrsson, H. H., Wiech, K., Weiskopf, N., Dolan, R. J., and Passingham, R. E. (2007). Threatening a Rubber Hand that You Feel is Yours Elicits a Cortical Anxiety Response. *Proc. Natl. Acad. Sci.* 104, 9828–9833. doi:10.1073/pnas.0610011104

Faller, J., Cummings, J., Saproo, S., and Sajda, P. (2019). Regulation of Arousal via Online Neurofeedback Improves Human Performance in a Demanding Sensory-Motor Task. *Proc. Natl. Acad. Sci. U.S.A.* 116, 6482–6490. doi:10.1073/pnas.1817207116

Felnhofer, A., Kothgassner, O. D., Hetterle, T., Beutl, L., Hlavacs, H., and Kryspin-Exner, I. (2014). Afraid to be There? Evaluating the Relation between Presence, Self-Reported Anxiety, and Heart Rate in a Virtual Public Speaking Task. *Cyberpsychol. Behav. Soc. Netw.* 17, 310–316. doi:10.1089/cyber.2013.0472

Felnhofer, A., Kothgassner, O. D., Schmidt, M., Heinzle, A.-K., Beutl, L., Hlavacs, H., et al. (2015). Is Virtual Reality Emotionally Arousing? Investigating Five Emotion Inducing Virtual Park Scenarios. *Int. J. Hum. Comput. Stud.* 82, 48–56. doi:10.1016/j.ijhcs.2015.05.004

Fernandez, J. A., Fusté, A., Richer, R., and Maes, P. (2019). "Deep Reality: An Underwater VR Experience to Promote Relaxation by Unconscious HR, EDA, and Brain Activity Biofeedback," in ACM SIGGRAPH 2019 Virtual, Augmented, and Mixed Reality (New York, NY: Association for Computing Machinery).

Fominykh, M., Prasolova-Førland, E., Stiles, T. C., Krogh, A. B., and Linde, M. (2018). Conceptual Framework for Therapeutic Training With Biofeedback in Virtual Reality: First Evaluation of a Relaxation Simulator. *J. Interact. Learn. Res.* 29, 51–75.

Freire, R. C., Ferreira-Garcia, R., Cabo, M. C., Martins, R. M., and Nardi, A. E. (2020). Panic Attack Provocation in Panic Disorder Patients with a Computer Simulation. *J. Affect. Disord.* 264, 498–505. doi:10.1016/j.jad.2019.11.081

Gaillard, A. (1993). Comparing the Concepts of Mental Load and Stress. *Ergonomics* 36, 991–1005. doi:10.1080/00140139308967972

Gamito, P., Oliveira, J., Baptista, A., Morais, D., Lopes, P., Rosa, P., et al. (2014). Eliciting Nicotine Craving with Virtual Smoking Cues. *Cyberpsychol. Behav. Soc. Netw.* 17, 556–561. doi:10.1089/cyber.2013.0329

Gao, J., Liu, S., Feng, Q., Zhang, X., Jiang, M., Wang, L., et al. (2019). Subjective and Objective Quantification of the Effect of Distraction on Physician's Workload and Performance During Simulated Laparoscopic Surgery. *Med. Sci. Monit.* 25, 3127–3132. doi:10.12659/MSM.914635

García-Rodríguez, O., Weidberg, S., Gutiérrez-Maldonado, J., and Secades-Villa, R. (2013). Smoking a Virtual Cigarette Increases Craving Among Smokers. *Addict. Behav.* 38, 2551–2554. doi:10.1016/j.addbeh.2013.05.007

Gavgani, A. M., Nesbitt, K. V., Blackmore, K. L., and Nalivaiko, E. (2017). Profiling Subjective Symptoms and Autonomic Changes Associated with Cybersickness. *Auton. Neurosci.* 203, 41–50. doi:10.1016/j.autneu.2016.12.004

Gersak, G., Lu, H., and Guna, J. (2020). Effect of VR Technology Matureness on VR Sickness. *Multimed. Tools. Appl.* 79, 14491–14507. doi:10.1007/s11042-018-6969-2

Gonzalez, D. S., Moro, A. D., Quintero, C., and Sarmiento, W. J. (2016). "Fear Levels in Virtual Environments, an Approach to Detection and Experimental User Stimuli Sensation," in 2016 XXI Symposium on Signal Processing, Images and Artificial Vision (STSIVA), 1–6.

González-Franco, M., Peck, T. C., Rodríguez-Fornells, A., and Slater, M. (2014). A Threat to a Virtual Hand Elicits Motor Cortex Activation. *Exp. Brain Res.* 232, 875–887. doi:10.1007/s00221-013-3800-1

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., et al. (2013). Meg and Eeg Data Analysis with mne-Python. *Front. Neurosci.* 7, 267. doi:10.3389/fnins.2013.00267

Granato, M., Gadia, D., Maggiorini, D., and Ripamonti, L. A. (2020). An Empirical Study of Players' Emotions in VR Racing Games Based on a Dataset of Physiological Data. *Multimed. Tools Appl.* 79, 33657–33686. doi:10.1007/s11042-019-08585-y

Greinacher, R., Kojić, T., Meier, L., Parameshappa, R. G., Möller, S., and Voigt-Antons, J. (2020). "Impact of Tactile and Visual Feedback on Breathing Rhythm and User Experience in VR Exergaming," in 2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX), 1–6.

Grimm, P. (2010). *Social Desirability Bias*. Hoboken, United States: Wiley.

Grübel, J., Weibel, R., Jiang, M. H., Hölscher, C., Hackman, D. A., and Schinazi, V. R. (2016). "Eve: A Framework for Experiments in Virtual Environments," in *Spatial Cognition X*. Bremen, Germany: Springer, 159–176.

Guna, J., Gersak, G., Humar, I., Krebl, M., Orel, M., Lu, H., et al. (2020). Virtual Reality Sickness and Challenges behind Different Technology and Content Settings. *Mob. Netw. Appl.* 25, 1436–1445. doi:10.1007/s11036-019-01373-w

Guna, J., Gersak, G., Humar, I., Song, J., Drnovsek, J., and Pogacnik, M. (2019). Influence of Video Content Type on Users' Virtual Reality Sickness Perception and Physiological Response. *Future Gener. Comput. Syst.* 91, 263–276. doi:10.1016/j.future.2018.08.049

Gupta, K., Hajika, R., Pai, Y. S., Duenser, A., Lochner, M., and Billinghurst, M. (2019). "AI We Trust: Investigating the Relationship between Biosignals, Trust and Cognitive Load in VR," in 25th ACM Symposium on Virtual Reality Software and TechnologyVRST '19, Parramatta, Australia (New York, NY: Association for Computing Machinery).

Gupta, K., Hajika, R., Pai, Y. S., Duenser, A., Lochner, M., and Billinghurst, M. (2020). "Measuring Human Trust in a Virtual Assistant Using Physiological Sensing in Virtual Reality," in 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), 756–765.

Haller, J. C., Jang, Y. H., Haller, J., Shaw, L., and Wünsche, B. C. (2019). "HIIT the Road: Using Virtual Spectator Feedback in HIIT-Based Exergaming," in Proceedings of the Australasian Computer Science Week Multiconference, Sydney, Australia (New York, NY: Association for Computing Machinery).

Ham, J., Cho, D., Oh, J., and Lee, B. (2017). "Discrimination of Multiple Stress Levels in Virtual Reality Environments Using Heart Rate variabilityIEEE Engineering in Medicine and Biology Society," in 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 3989–3992.

Hancock, P. A., Meshkati, N., and Robertson, M. (1985). Physiological Reflections of Mental Workload. *Aviat. Space Environ. Med.* 56, 1110–1114.

Hart, S. G., and Staveland, L. E. (1988). Development of Nasa-Tlx (Task Load Index): Results of Empirical and Theoretical Research. *Adv. Psychol.* 52, 139–183. doi:10.1016/s0166-4115(08)62386-9

Herborn, K. A., Graves, J. L., Jerem, P., Evans, N. P., Nager, R., McCafferty, D. J., et al. (2015). Skin Temperature Reveals the Intensity of Acute Stress. *Physiol. Behav.* 152, 225–230. doi:10.1016/j.physbeh.2015.09.032

Herumurti, D., Yuniarti, A., Rimawan, P., and Yunanto, A. A. (2019). "Overcoming Glossophobia Based on Virtual Reality and Heart Rate Sensors," in 2019 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT), 139–144.

Hildebrandt, L. K., McCall, C., Engen, H. G., and Singer, T. (2016). Cognitive Flexibility, Heart Rate Variability, and Resilience Predict fine-grained Regulation of Arousal during Prolonged Threat. *Psychophysiology* 53, 880–890. doi:10.1111/psyp.12632

Hofmann, S. M., Klotzsche, F., Mariola, A., Nikulin, V. V., Villringer, A., and Gaebler, M. (2018). "Decoding Subjective Emotional Arousal during a Naturalistic VR Experience from EEG Using LSTMs," in 2018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR), 128–131.

Hogervorst, M. A., Brouwer, A.-M., and Van Erp, J. B. (2014). Combining and Comparing Eeg, Peripheral Physiology and Eye-Related Measures for the Assessment of Mental Workload. *Front. Neurosci.* 8, 322. doi:10.3389/fnins.2014.00322

Houzangbe, S., Christmann, O., Gorisse, G., and Richir, S. (2019). "Effects of Voluntary Heart Rate Control on User Engagement in Virtual Reality," in 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), 982–983.

Houzangbe, S., Christmann, O., Gorisse, G., and Richir, S. (2018). "Fear as a Biofeedback Game Mechanic in Virtual Reality: Effects on Engagement and Perceived Usability," in Proceedings of the 13th International Conference on the Foundations of Digital Games FDG '18, Malmö, Sweden (New York, NY: Association for Computing Machinery).

Hoxhallari, E., Behr, I. J., Bradshaw, J. S., Morkos, M. S., Haan, P. S., Schaefer, M. C., et al. (2019). Virtual Reality Improves the Patient Experience during Wide-Awake Local Anesthesia No Tourniquet Hand Surgery: A Single-Blind, Randomized, Prospective Study. *Plast. Reconstr. Surg.* 144, 408–414. doi:10.1097/PRS.0000000000005831

Hu, F., Wang, H., Chen, J., and Gong, J. (2018). "Research on the Characteristics of Acrophobia in Virtual Altitude Environment," in 2018 IEEE International Conference on Intelligence and Safety for Robotics (ISR), 238–243.

Ishaque, S., Rueda, A., Nguyen, B., Khan, N., and Krishnan, S. (2020). "Physiological Signal Analysis and Classification of Stress from Virtual Reality Video Game," in 2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC), 867–870.

Jänig, W. (2008). *Integrative Action of the Autonomic Nervous System: Neurobiology of Homeostasis*. Cambridge, UK: Cambridge University Press.

Jeong, D., Yoo, S., and Yun, J. (2019). "Cybersickness Analysis with EEG Using Deep Learning Algorithms," in 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), 827–835.

John, B. (2019). "Pupil Diameter as a Measure of Emotion and Sickness in VR," in Proceedings of the 11th ACM Symposium on Eye Tracking Research & ApplicationsETRA '19, Denver, CO (New York, NY: Association for Computing Machinery).

Kahlon, S., Lindner, P., and Nordgreen, T. (2019). Virtual Reality Exposure Therapy for Adolescents with Fear of Public Speaking: a Non-randomized Feasibility and Pilot Study. *Child. Adolesc. Psychiatry Ment. Health* 13, 47. doi:10.1186/s13034-019-0307-y

Kakkos, I., Dimitrakopoulos, G. N., Gao, L., Zhang, Y., Qi, P., Matsopoulos, G. K., et al. (2019). Mental Workload Drives Different Reorganizations of Functional Cortical Connectivity between 2D and 3D Simulated Flight Experiments. *IEEE Trans. Neural Syst. Rehabil. Eng.* 27, 1704–1713. doi:10.1109/TNSRE.2019.2930082

Kaminskas, V., and Sciglinskas, E. (2019). A Comparison of the Control Schemes of Human Response to a Dynamic Virtual 3D Face. *Inf. Technol. Control* 48, 250–267. doi:10.5755/j01.itc.48.2.21667

Kataoka, H., Kano, H., Yoshida, H., Saijo, A., Yasuda, M., and Osumi, M. (1998). "Development of a Skin Temperature Measuring System for Non-contact Stress Evaluation," in Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE), 940–943.

Kaur, R., Sun, R., Ziegelman, L., Sowers, R., and Hernandez, M. E. (2019). "Using Virtual Reality to Examine the Neural and Physiological Anxiety-Related Responses to Balance-Demanding Target-Reaching Leaning Tasks," in 2019 IEEE-RAS 19th International Conference on Humanoid Robots (Humanoids), 1–7.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., and Lilienthal, M. G. (1993). Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *Int. J. Aviat. Psychol.* 3, 203–220. doi:10.1207/s15327108ijap0303_3

Kerous, B., Bartecek, R., Roman, R., Sojka, P., Becev, O., and Liarokapis, F. (2020). Examination of Electrodermal and Cardio-Vascular Reactivity in Virtual Reality through a Combined Stress Induction Protocol. *J. Ambient. Intell. Humaniz. Comput.* 11, 6033–6042. doi:10.1007/s12652-020-01858-7

Khokhar, A., Yoshimura, A., and Borst, C. W. (2019). "Pedagogical Agent Responsive to Eye Tracking in Educational VR," in 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), 1018–1019.

Kim, H.-G., Cheon, E.-J., Bai, D.-S., Lee, Y. H., and Koo, B.-H. (2018). Stress and Heart Rate Variability: A Meta-Analysis and Review of the Literature. *Psychiatry Invest.* 15, 235. doi:10.30773/pi.2017.08.17

Kirsch, K., Schatzschneider, C., Garber, C., Rosenberger, A., Kirsten, K., Ariza, O., et al. (2019). "KiVR Sports: Influencing the Users Physical Activity in VR by Using Audiovisual Stimuli in Exergames," in Proceedings of Mensch Und Computer 2019, MuC'19, Hamburg, Germany (New York, NY: Association for Computing Machinery), 777–781.

Kishimoto, T., and Ding, X. (2019). The Influences of Virtual Social Feedback on Social Anxiety Disorders. *Behav. Cogn. Psychother.* 47, 726–735. doi:10.1017/S1352465819000377

Kivelä, O., Alavesa, P., Visuri, A., and Ojala, T. (2019). "Study on the Motivational and Physical Effects of Two VR Exergames," in 2019 11th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games), 1–2.

Kocur, M., Dechant, M. J., Lankes, M., Wolff, C., and Mandryk, R. (2020). "Eye Caramba: Gaze-Based Assistance for Virtual Reality Aiming and Throwing Tasks in Games," in ACM Symposium on Eye Tracking Research and Applications, ETRA '20, Stuttgart, Germany (New York, NY: Association for Computing Machinery).

Kojić, T., Nugyen, L. T., and Voigt-Antons, J. (2019). "Impact of Constant Visual Biofeedback on User Experience in Virtual Reality Exergames," in 2019 IEEE International Symposium on Multimedia (ISM), 307–3073.

Kokkinara, E., Kilteni, K., Blom, K. J., and Slater, M. (2016). First Person Perspective of Seated Participants Over a Walking Virtual Body Leads to Illusory Agency Over the Walking. *Sci. Rep.* 6, 1–11. doi:10.1038/srep28879

Kothgassner, O. D., Felnhofer, A., Hlavacs, H., Beutl, L., Palme, R., Kryspin-Exner, I., et al. (2016). Salivary Cortisol and Cardiovascular Reactivity to a Public Speaking Task in a Virtual and Real-Life Environment. *Comput. Hum. Behav.* 62, 124–135. doi:10.1016/j.chb.2016.03.081

Kothgassner, O. D., Goreis, A., Kafka, J. X., Kaufmann, M., Atteneder, K., Beutl, L., et al. (2019). Virtual Social Support Buffers Stress Response: An Experimental Comparison of Real-Life and Virtual Support Prior to a Social Stressor. *J. Behav. Ther. Exp. Psychiatry* 63, 57–65. doi:10.1016/j.jbtep.2018.11.003

Koticha, P., Katge, F., Shetty, S., and Patil, D. P. (2019). Effectiveness of Virtual Reality Eyeglasses as a Distraction Aid to Reduce Anxiety Among 6-10-Year-Old Children Undergoing Dental Extraction Procedure. *Int. J. Clin. Pediatr. Dent.* 12, 297–302. doi:10.5005/jp-journals-10005-1640

Kreibig, S. D. (2010). Autonomic Nervous System Activity in Emotion: A Review. *Biol. Psychol.* 84, 394–421. doi:10.1016/j.biopsycho.2010.03.010

Krogmeier, C., Mousas, C., and Whittinghill, D. (2019). Human-Virtual Character Interaction: Toward Understanding the Influence of Haptic Feedback. *Comput. Animat. Virtual Worlds* 30, 1883. doi:10.1002/cav.1883

Kuriakose, S., Kunche, S., Narendranath, B., Jain, P., Sonker, S., and Lahiri, U. (2013). "A Step towards Virtual Reality Based Social Communication for Children with Autism," in 2013 International Conference on Control, Automation, Robotics and Embedded Systems (CARE) (IEEE), 1–6.

Kurniawan, H., Maslov, A. V., and Pechenizkiy, M. (2013). "Stress Detection from Speech and Galvanic Skin Response Signals," in Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems (IEEE), 209–214.

Lai, C. Q., Ibrahim, H., Abdullah, M. Z., Abdullah, J. M., Suandi, S. A., and Azman, A. (2018). "Literature Survey on Applications of Electroencephalography (EEG)," in AIP Conference Proceedings (AIP Publishing LLC), 020070.

Laight, D. (2013). Overview of Peripheral Nervous System Pharmacology. *Nurse Prescribing* 11, 448–454. doi:10.12968/npre.2013.11.9.448

Latoschik, M. E., and Wienrich, C. (2021). Coherence and Plausibility, Not Presence?! Pivotal Conditions for Xr Experiences and Effects, a Novel Model. *arXiv*.

LaViola, J. J., Jr (2000). A Discussion of Cybersickness in Virtual Environments. *ACM Sigchi Bull.* 32, 47–56. doi:10.1145/333329.333344

Lee, S.-H., Kim, Y.-M., and Lee, B.-H. (2015). Effects of Virtual Reality-Based Bilateral Upper-Extremity Training on Brain Activity in Post-Stroke Patients. *J. Phys. Ther. Sci.* 27, 2285–2287. doi:10.1589/jpts.27.2285

Levy, F., Leboucher, P., Rautureau, G., and Jouvent, R. (2016). E-virtual Reality Exposure Therapy in Acrophobia: A Pilot Study. *J. Telemed. Telecare* 22, 215–220. doi:10.1177/1357633X15598243

Li, Y., Chiu, P., Yeh, S., and Zhou, C. (2017). "Effects of Virtual Reality and Augmented Reality on Induced Anxiety," in 2017 5th International Conference on Enterprise Systems (ES), 132–138.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., et al. (2009). The Prisma Statement for Reporting Systematic Reviews and Meta-Analyses of Studies that Evaluate Health Care Interventions: Explanation and Elaboration. *J. Clin. Epidemiol.* 62, e1–e34. doi:10.1016/j.jclinepi.2009.06.006

Líndal, P. J., Jóhannsdóttir, K. R., Kristjánsson, U., Lensing, N., Stühmeier, A., Wohlan, A., et al. (2018). "Comparison of Teleportation and Fixed Track Driving in VR," in 2018 10th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games), 1–7.

Liszio, S., Graf, L., and Masuch, M. (2018). The Relaxing Effect of Virtual Nature: Immersive Technology Provides Relief in Acute Stress Situations. *Annu. Rev. CyberTherapy Telemed.* 16, 87–93.

Liu, Y., Lin, Y., Wu, S., Chuang, C., Prasad, M., and Lin, C. (2014). "EEG-Based Driving Fatigue Prediction System Using Functional-Link-Based Fuzzy Neural Network," in 2014 International Joint Conference on Neural Networks (IJCNN), 4109–4113.

Loreto, C. D., Chardonnet, J., Ryard, J., and Rousseau, A. (2018). "WoaH: A Virtual Reality Work-At-Height Simulator," in 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), 281–288.

Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., et al. (2018). A Review of Classification Algorithms for eeg-Based Brain–Computer Interfaces: A 10 Year Update. *J. Neural Eng.* 15, 031005. doi:10.1088/1741-2552/aab2f2

Lou, J., Wang, Y., Nduka, C., Hamedi, M., Mavridou, I., Wang, F.-Y., et al. (2020). Realistic Facial Expression Reconstruction for VR HMD Users. *IEEE Trans. Multimedia* 22, 730–743. doi:10.1109/TMM.2019.2933338

Luck, M., and Aylett, R. (2000). Applying Artificial Intelligence to Virtual Reality: Intelligent Virtual Environments. *Appl. Artif. Intell.* 14, 3–32. doi:10.1080/088395100117142

Lugrin, J.-L., Ertl, M., Krop, P., Klüpfel, R., Stierstorfer, S., Weisz, B., et al. (2018). "Any "Body" There? - Avatar Visibility Effects in a Virtual Reality Game," in 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), 17–24.

Lugrin, J.-L., Latoschik, M. E., Glémarec, Y., Bosser, A.-G., Chollet, M., and Lugrin, B. (2019a). "Towards Narrative-Driven Atmosphere for Virtual Classroom," in Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, 1–6.

Lugrin, J.-L., Latt, J., and Latoschik, M. E. (2015). "Avatar Anthropomorphism and Illusion of Body Ownership in vr," in 2015 IEEE Virtual Reality (VR) (IEEE), 229–230.

Lugrin, J.-L., Unruh, F., Landeck, M., Lamour, Y., Latoschik, M. E., Vogeley, K., et al. (2019b). "Experiencing Waiting Time in Virtual Reality," in Proceedings of the 25th ACM Conference on Virtual Reality Software and Technology.

Ma, K., and Hommel, B. (2015). The Role of agency for Perceived Ownership in the Virtual Hand Illusion. *Conscious. Cogn.* 36, 277–288. doi:10.1016/j.concog.2015.07.008

Ma, L., Zhao, X., Li, Z., Zhao, M., and Xu, Z. (2018). "A sEMG-Based Hand Function Rehabilitation System for Stroke Patients," in 2018 3rd International Conference on Advanced Robotics and Mechatronics (ICARM), 497–502.

Macauda, G., Bertolini, G., Palla, A., Straumann, D., Brugger, P., and Lenggenhager, B. (2015). Binding Body and Self in Visuo-Vestibular Conflicts. *Eur. J. Neurosci.* 41, 810–817. doi:10.1111/ejn.12809

Malta, L. S., Giosan, C., Szkodny, L. E., Altemus, M. M., Rizzo, A. A., Silbersweig, D. A., et al. (2020). Development of a Virtual Reality Laboratory Stressor. *Virtual Reality*. doi:10.1007/s10055-020-00455-5

Maples-Keller, J. L., Rauch, S. A. M., Jovanovic, T., Yasinski, C. W., Goodnight, J. M., Sherrill, A., et al. (2019). Changes in Trauma-Potentiated Startle, Skin Conductance, and Heart Rate within Prolonged Exposure Therapy for PTSD in High and Low Treatment Responders. *J. Anxiety Disord.* 68, 102147. doi:10.1016/j.janxdis.2019.102147

Marín-Morales, J., Higuera-Trujillo, J. L., Greco, A., Guixeres, J., Llinares, C., Scilingo, E. P., et al. (2018). Affective Computing in Virtual Reality: Emotion Recognition from Brain and Heartbeat Dynamics Using Wearable Sensors. *Sci. Rep.* 8, 13657. doi:10.1038/s41598-018-32063-4

Marín-Morales, J., Llinares, C., Guixeres, J., and Alcañiz, M. (2020). Emotion Recognition in Immersive Virtual Reality: From Statistics to Affective Computing. *Sensors* 20, 5163. doi:10.3390/s20185163

Masaoka, Y., and Homma, I. (1997). Anxiety and Respiratory Patterns: Their Relationship during Mental Stress and Physical Load. *Int. J. Psychophysiol.* 27, 153–159. doi:10.1016/s0167-8760(97)00052-4

Mavridou, I., Seiss, E., Hamedi, M., Balaguer-Ballester, E., and Nduka, C. (2018a). "Towards Valence Detection from emg for Virtual Reality Applications," in 12th International Conference on Disability, Virtual Reality and Associated

Technologies (ICDVRAT 2018), Nottingham, United Kingdom, September 4–6, 2018.

Mavridou, I., Seiss, E., Kostoulas, T., Nduka, C., and Balaguer-Ballester, E. (2018b). "Towards an Effective Arousal Detection System for Virtual Reality," in Proceedings of the Workshop on Human-Habitat for Health (H3): Human-Habitat Multimodal Interaction for Promoting Health and Well-Being in the Internet of Things Era, H3 '18, Boulder, CO (New York, NY: Association for Computing Machinery).

McDonough, D. J., Pope, Z. C., Zeng, N., Liu, W., and Gao, Z. (2020). Comparison of College Students' Blood Pressure, Perceived Exertion, and Psychosocial Outcomes During Virtual Reality, Exergaming, and Traditional Exercise: An Exploratory Study. *Games Health J.* 9, 290–296. doi:10.1089/g4h.2019.0196

Melillo, P., Bracale, M., and Pecchia, L. (2011). Nonlinear Heart Rate Variability Features for Real-Life Stress Detection. Case Study: Students Under Stress Due to university Examination. *Biomed. Eng. Online* 10, 96. doi:10.1186/1475-925x-10-96

Melnyk, R., Campbell, T., Holler, T., Cameron, K., Saba, P., Witthaus, M., et al. (2021). See Like an Expert: Gaze-Augmented Training Enhances Skill Acquisition in a Virtual Reality Robotic Suturing Task. *J. Endourol.* 35, 376–382. doi:10.1089/end.2020.0445

Mertens, G., Wagensveld, P., and Engelhard, I. M. (2019). Cue Conditioning Using a Virtual Spider Discriminates Between High and Low Spider Fearful Individuals. *Comput. Hum. Behav.* 91, 192–200. doi:10.1016/j.chb.2018.10.006

Mishra, N., and Folmer, E. (2018). "Measuring Physical Exertion in Virtual Reality Exercise Games," in Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology, Tokyo, Japan (New York, NY: Association for Computing Machinery).

Moon, S.-E., and Lee, J.-S. (2016). Implicit Analysis of Perceptual Multimedia Experience Based on Physiological Response: A Review. *IEEE Trans. Multimed.* 19, 340–353. doi:10.1109/TMM.2016.2614880

Morii, M., Sakagami, T., Masuda, S., Okubo, S., and Tamari, Y. (2017). How Does Response Bias Emerge in Lengthy Sequential Preference Judgments?. *Behaviormetrika* 44, 575–591. doi:10.1007/s41237-017-0036-6

Mosquera, C., Galvan, B., Liu, E., De Vito, R., Ting, P., Costello, E. L., et al. (2019). "ANX Dread: A Virtual Reality Experience to Explore Anxiety during Task Completion," in Proceedings of the 14th International Conference on the Foundations of Digital Games, FDG '19, San Luis Obispo, CA (New York, NY: Association for Computing Machinery).

Mostajeran, F., Balci, M. B., Steinicke, F., Kühn, S., and Gallinat, J. (2020). "The Effects of Virtual Audience Size on Social Anxiety during Public Speaking," in 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), 303–312.

Mueller, E., Stolz, C., and Endres, D. (2017). Threat-Conditioned Contexts Modulate the Late Positive Potential to Neutral vs. Angry Avatars - A Mobile-EEG/Virtual Reality Study. *Psychophysiology* 54, S2. doi:10.1111/psyp.12921

Muñoz, J. E., Paulino, T., Vasanth, H., and Baras, K. (2016). "Physiovr: A Novel mobile Virtual Reality Framework for Physiological Computing," in 2016 IEEE 18th international conference on e-Health Networking, Applications and Services (Healthcom) (IEEE), 1–6.

Murray, E. G., Neumann, D. L., Moffitt, R. L., and Thomas, P. R. (2016). The Effects of the Presence of Others During a Rowing Exercise in a Virtual Reality Environment. *Psychol. Sport Exerc.* 22, 328–336. doi:10.1016/j.psychsport.2015.09.007

Ong, T. L., Ruppert, M. M., Akbar, M., Rashidi, P., Ozrazgat-Baslanti, T., Bihorac, A., et al. (2020). Improving the Intensive Care Patient Experience With Virtual Reality-A Feasibility Study. *Crit. Care Explor.* 2, e0122. doi:10.1097/CCE.0000000000000122

Orlosky, J., Huynh, B., and Hollerer, T. (2019). "Using Eye Tracked Virtual Reality to Classify Understanding of Vocabulary in Recall Tasks," in 2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR), 66–667.

Parenthoen, M., Murie, F., and Thery, F. (2015). "The Sea Is Your Mirror," in Proceedings of the 8th ACM SIGGRAPH Conference on Motion in Games, MIG '15, Paris, France (New York, NY: Association for Computing Machinery), 159–165.

Park, S. K., Yang, D. J., Uhm, Y. H., Heo, J. W., and Kim, J. H. (2016). The Effect of Virtual Reality-Based Eccentric Training on Lower Extremity Muscle Activation and Balance in Stroke Patients. *J. Phys. Ther. Sci.* 28, 2055–2058. doi:10.1589/jpts.28.2055

Patel, J., Fluet, G., Merians, A., Qiu, Q., Yarossi, M., Adamovich, S., et al. (2015). "Virtual Reality-Augmented Rehabilitation in the Acute Phase post-stroke for Individuals with Flaccid Upper Extremities: A Feasibility Study," in 2015 International Conference on Virtual Rehabilitation (ICVR), 215–223.

Patel, J., Qiu, Q., Yarossi, M., Merians, A., Massood, S., Tunik, E., et al. (2017). Exploring the Impact of Visual and Movement Based Priming on a Motor Intervention in the Acute Phase post-stroke in Persons with Severe Hemiparesis of the Upper Extremity. *Disabil. Rehabil.* 39, 1515–1523. doi:10.1080/09638288.2016.1226419

Patibanda, R., Mueller, F. F., Leskovsek, M., and Duckworth, J. (2017). "Life Tree: Understanding the Design of Breathing Exercise Games," in Proceedings of the Annual Symposium on Computer-Human Interaction in Play, Amsterdam, Netherlands (New York, NY: Association for Computing Machinery), 19–31.

Plouzeau, J., Chardonnet, J., and Merienne, F. (2018). "Using Cybersickness Indicators to Adapt Navigation in Virtual Reality: A Pre-Study," in 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), 661–662. doi:10.1109/VR.2018.8446192

Posner, J., Russell, J. A., and Peterson, B. S. (2005). The Circumplex Model of Affect: An Integrative Approach to Affective Neuroscience, Cognitive Development, and Psychopathology. *Dev. Psychopathol.* 17, 715. doi:10.1017/S0954579405050340

Prachyabrued, M., Wattanadhirach, D., Dudrow, R. B., Krairojananan, N., and Fuengfoo, P. (2019). "Toward Virtual Stress Inoculation Training of Prehospital Healthcare Personnel: A Stress-Inducing Environment Design and Investigation of an Emotional Connection Factor," in 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), 671–679.

Preuss, N., and Ehrsson, H. H. (2019). Full-Body Ownership Illusion Elicited by Visuo-Vestibular Integration. *J. Exp. Psychol. Hum. Percept. Perform.* 45, 209–223. doi:10.1037/xhp0000597

Pullman, S. L., Goodin, D. S., Marquinez, A. I., Tabbal, S., and Rubin, M. (2000). Clinical Utility of Surface Emg: Report of the Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology. *Neurology* 55, 171–177. doi:10.1212/wnl.55.2.171

Quintero, L., Papapetrou, P., and Muñoz, J. E. (2019). "Open-source Physiological Computing Framework Using Heart Rate Variability in mobile Virtual Reality Applications," in 2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR) (IEEE), 126–1267.

Rahman, Y., Asish, S. M., Fisher, N. P., Bruce, E. C., Kulshreshth, A. K., and Borst, C. W. (2020). "Exploring Eye Gaze Visualization Techniques for Identifying Distracted Students in Educational VR," in 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), 868–877.

Ramdhani, N., Akpewila, F., Faizah, M., and Resibisma, B. (2019). "It's So Real: Psychophysiological Reaction Towards Virtual Reality Exposure," in 2019 5th International Conference on Science and Technology (ICST), 1–5.

Rangelova, S., Flutura, S., Huber, T., Motus, D., and André, E. (2019). "Exploration of Physiological Signals Using Different Locomotion Techniques in a vr Adventure Game," in International Conference on Human-Computer Interaction (Springer), 601–616.

Rao, D. G., Havale, R., Nagaraj, M., Karobari, N. M., Latha, A. M., Tharay, N., et al. (2019). Assessment of Efficacy of Virtual Reality Distraction in Reducing Pain Perception and Anxiety in Children Aged 6-10 Years: A Behavioral Interventional Study. *Int. J. Clin. Pediatr. Dent.* 12, 510–513. doi:10.5005/jp-journals-10005-1694

Ravaja, N., Bente, G., Kätsyri, J., Salminen, M., and Takala, T. (2018). Virtual Character Facial Expressions Influence Human Brain and Facial EMG Activity in a Decision-Making Game. *IEEE Trans. Affect. Comput.* 9, 285–298. doi:10.1109/TAFFC.2016.2601101

Rebenitsch, L., and Owen, C. (2016). Review on Cybersickness in Applications and Visual Displays. *Virtual Real.* 20, 101–125. doi:10.1007/s10055-016-0285-9

Redline, C. D., Dillman, D. A., Carley-Baxter, L., and Creecy, R. (2003). "Factors that Influence reading and Comprehension in Self-Administered Questionnaires," in Workshop on Item-Nonresponse and Data Quality, Basel Switzerland.

Richard, M.-O., and Toffoli, R. (2009). Language Influence in Responses to Questionnaires by Bilingual Respondents: A Test of the Whorfian Hypothesis. *J. Bus. Res.* 62, 987–994. doi:10.1016/j.jbusres.2008.10.016

Robertson, A., Khan, R., Fick, D., Robertson, W. B., Gunaratne, D. R., Yapa, S., et al. (2017). "The Effect of Virtual Reality in Reducing Preoperative Anxiety in Patients Prior to Arthroscopic Knee Surgery: A Randomised Controlled Trial," in 2017 IEEE 5th International Conference on Serious Games and Applications for Health (SeGAH), 1–7.

Robins, R. W., Fraley, R. C., and Krueger, R. F. (2009). *Handbook of Research Methods in Personality Psychology*. New York, USA: Guilford Press.

Robitaille, P., and McGuffin, M. J. (2019). "Increased Affect-Arousal in VR Can Be Detected from Faster Body Motion with Increased Heart Rate," in Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games, I3D '19, Montreal, Canada (New York, NY: Association for Computing Machinery).

Roth, D., and Latoschik, M. E. (2020). Construction of the Virtual Embodiment Questionnaire (Veq). *IEEE Ann. Hist. Comput.* 26, 3546–3556. doi:10.1109/tvcg.2020.3023603

Roth, D., Lugrin, J.-L., Latoschik, M. E., and Huber, S. (2017). "Alpha Ivbo - Construction of a Scale to Measure the Illusion of Virtual Body Ownership," in Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, 2875–2883.

Rowley, J. (2014). Designing and Using Research Questionnaires. *Manage. Res. Rev.* 37, 308–330.

Russell, J. A., and Mehrabian, A. (1977). Evidence for a Three-Factor Theory of Emotions. *J. Res. Personal.* 11, 273–294. doi:10.1016/0092-6566(77)90037-x

Saha, D. P., Martin, L. T., and Knapp, R. B. (2018). "Towards Defining a Quality-Metric for Affective Feedback in an Intelligent Environment," in 2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops) (IEEE), 609–614.

Sakamoto, K., Shirai, S., Orlosky, J., Nagataki, H., Takemura, N., Alizadeh, M., et al. (2020). "Exploring Pupillometry as a Method to Evaluate Reading Comprehension in VR-Based Educational Comics," in 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), 422–426.

Salkevicius, J., Damasevicius, R., Maskeliunas, R., and Laukiene, I. (2019). Anxiety Level Recognition for Virtual Reality Therapy System Using Physiological Signals. *Electronics* 8. doi:10.3390/electronics8091039

Salminen, M., Járvelá, S., Ruonala, A., Harjunen, V., Jacucci, G., Hamari, J., et al. (2019). Evoking Physiological Synchrony and Empathy Using Social VR with Biofeedback. *IEEE Trans. Affect. Comput.*. doi:10.1109/TAFFC.2019.2958657

Samson, A. C., Kreibig, S. D., Soderstrom, B., Wade, A. A., and Gross, J. J. (2016). Eliciting Positive, Negative and Mixed Emotional States: A Film Library for Affective Scientists. *Cogn. Emot.* 30, 827–856. doi:10.1080/02699931.2015.1031089

Schroeder, R. (2010). *Being There Together: Social Interaction in Shared Virtual Environments*. Oxford, UK: Oxford University Press.

Schwind, V., Knierim, P., Haas, N., and Henze, N. (2019). "Using Presence Questionnaires in Virtual Reality," in Proceedings of the 2019 CHI conference on human factors in computing systems, 1–12.

Shiban, Y., Diemer, J., Brandl, S., Zack, R., Mühlberger, A., and Wüst, S. (2016a). Trier Social Stress Test *In Vivo* and in Virtual Reality: Dissociation of Response Domains. *Int. J. Psychophysiol.* 110, 47–55. doi:10.1016/j.ijpsycho.2016.10.008

Shiban, Y., Diemer, J., Müller, J., Brütting-Schick, J., Pauli, P., and Mühlberger, A. (2017). Diaphragmatic Breathing During Virtual Reality Exposure Therapy for Aviophobia: Functional Coping Strategy or Avoidance Behavior? A Pilot Study. *BMC Psychiatry* 17, 29. doi:10.1186/s12888-016-1181-2

Shiban, Y., Peperkorn, H., Alpers, G. W., Pauli, P., and Mühlberger, A. (2016b). Influence of Perceptual Cues and Conceptual Information on the Activation and Reduction of Claustrophobic Fear. *J. Behav. Ther. Exp. Psychiatry* 51, 19–26. doi:10.1016/j.jbtep.2015.11.002

Shumailov, I., and Gunes, H. (2017). "Computational Analysis of Valence and Arousal in Virtual Reality Gaming Using Lower Arm Electromyograms," in 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII), 164–169.

Simões, M., Bernardes, M., Barros, F., and Castelo-Branco, M. (2018). Virtual Travel Training for Autism Spectrum Disorder: Proof-Of-Concept Interventional Study. *JMIR Serious Games* 6, e5. doi:10.2196/games.8428

Singh, H., and Singh, J. (2012). Human Eye Tracking and Related Issues: A Review. *Int. J. Sci. Res. Pub.* 2, 1–9. doi:10.5296/ijhrs.v2i3.2217

Siravenha, A. C., Reis, M. N., Cordeiro, I., Tourinho, R. A., Gomes, B. D., and Carvalho, S. R. (2019). "Residual MLP Network for Mental Fatigue Classification in Mining Workers from Brain Data," in 2019 8th Brazilian Conference on Intelligent Systems (BRACIS) (IEEE), 407–412.

Sirois, S., and Brisson, J. (2014). Pupillometry. *Wiley Interdiscip. Rev. Cogn. Sci.* 5, 679–692. doi:10.1002/wcs.1323

Skola, F., and Liarokapis, F. (2016). Examining the Effect of Body Ownership in Immersive Virtual and Augmented Reality Environments. *Vis. Comput.* 32, 761–770. doi:10.1007/s00371-016-1246-8

Slater, M., Pérez Marcos, D., Ehrsson, H., and Sanchez-Vives, M. V. (2008). Towards a Digital Body: The Virtual Arm Illusion. *Front. Hum. Neurosci.* 2, 6. doi:10.3389/neuro.09.006.2008

Slater, M., Usoh, M., and Steed, A. (1994). Depth of Presence in Virtual Environments. *Presence: Teleoperators Virtual Environ.* 3, 130–144. doi:10.1162/pres.1994.3.2.130

Solanki, D., and Lahiri, U. (2020). Adaptive Treadmill-Assisted Virtual Reality-Based Gait Rehabilitation for Post-Stroke Physical Reconditioning & Feasibility Study in Low-Resource Settings. *IEEE Access* 8, 88830–88843. doi:10.1109/ACCESS.2020.2994081

Soyka, F., Leyrer, M., Smallwood, J., Ferguson, C., Riecke, B. E., and Mohler, B. J. (2016). "Enhancing Stress Management Techniques Using Virtual Reality," in Proceedings of the ACM Symposium on Applied Perception, SAP '16, Anaheim, CA (New York, NY: Association for Computing Machinery), 85–88.

Spangler, D., Alam, S., Rahman, S., Crone, J., Robucci, R., Banerjee, N., et al. (2020). Multilevel Longitudinal Analysis of Shooting Performance as a Function of Stress and Cardiovascular Responses. *IEEE Trans. Affect. Comput.*. doi:10.1109/TAFFC.2020.2995769

Stauffert, J.-P., Niebling, F., and Latoschik, M. E. (2018). "Effects of Latency Jitter on Simulator Sickness in a Search Task," in 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR) (IEEE), 121–127.

Stauffert, J.-P., Niebling, F., and Latoschik, M. E. (2020). Latency and Cybersickness: Impact, Causes and Measures: A Review. *Front. Virtual Real.* 1, 31. doi:10.3389/frvir.2020.582204

Streck, A., Stepnicka, P., Klaubert, J., and Wolbers, T. (2019). "Neomento SAD - VR Treatment for Social Anxiety," in 2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR), 245–2451.

Suhaimi, N. S. B., Mountstephens, J., and Teo, J. (2020). "Emotional State Classification with Distributed Random Forest, Gradient Boosting Machine and Naïve Bayes in Virtual Reality Using Wearable Electroencephalography and Inertial Sensing," in 2020 IEEE 10th Symposium on Computer Applications Industrial Electronics (ISCAIE), 12–17.

Swidrak, J., and Pochwatko, G. (2019). "Being Touched by a Virtual Human.: Relationships between Heart Rate, Gender, Social Status, and Compliance," in Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents IVA '19, Paris, France (New York, NY: Association for Computing Machinery), 49–55.

Syrjamaki, A. H., Isokoski, P., Surakka, V., Pasanen, T. P., and Hietanen, J. K. (2020). Eye Contact in Virtual Reality - A Psychophysiological Study. *Comput. Hum. Behav.* 112, 106454. doi:10.1016/j.chb.2020.106454

Szczurowski, K., and Smith, M. (2017). "Measuring Presence: Hypothetical Quantitative Framework," in 2017 23rd International Conference on Virtual System Multimedia (VSMM), 1–8.

Tartarisco, G., Carbonaro, N., Tonacci, A., Bernava, G. M., Arnao, A., Crifaci, G., et al. (2015). Neuro-Fuzzy Physiological Computing to Assess Stress Levels in Virtual Reality Therapy. *Interact. Comput.* 27, 521–533. doi:10.1093/iwc/iwv010

Tarvainen, M. P., Niskanen, J.-P., Lipponen, J. A., Ranta-Aho, P. O., and Karjalainen, P. A. (2014). Kubios HRV–Heart Rate Variability Analysis Software. *Comp. Methods Programs Biomed.* 113, 210–220. doi:10.1016/j.cmpb.2013.07.024

Taylor, S., Jaques, N., Chen, W., Fedor, S., Sano, A., and Picard, R. (2015). "Automatic Identification of Artifacts in Electrodermal Activity Data," in 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (IEEE), 1934–1937.

Teo, J., and Chia, J. T. (2018). "Deep Neural Classifiers for Eeg-Based Emotion Recognition in Immersive Environments," in 2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE), 1–6.

Thompson-Lake, D. G. Y., Cooper, K. N., Mahoney, J. J., Bordnick, P. S., Salas, R., Kosten, T. R., et al. (2015). Withdrawal Symptoms and Nicotine Dependence Severity Predict Virtual Reality Craving in Cigarette-Deprived Smokers. *Nicotine Tob. Res.* 17, 796–802. doi:10.1093/ntr/ntu245

Tieri, G., Gioia, A., Scandola, M., Pavone, E. F., and Aglioti, S. M. (2017). Visual Appearance of a Virtual Upper Limb Modulates the Temperature of the Real Hand: a thermal Imaging Study in Immersive Virtual Reality. *Eur. J. Neurosci.* 45, 1141–1151. doi:10.1111/ejn.13545

Topalovic, U., Aghajan, Z. M., Villaroman, D., Hiller, S., Christov-Moore, L., Wishard, T. J., et al. (2020). Wireless Programmable Recording and Stimulation of Deep Brain Activity in Freely Moving Humans. *Neuron* 108, 322–334. doi:10.1016/j.neuron.2020.08.021

Tremmel, C. (2020). Estimating Cognitive Workload in an Interactive Virtual Reality Environment Using Electrophysiological and Kinematic Activity. PhD dissertation. (Norfolk, USA: Dominion University).

Tremmel, C., Herff, C., Sato, T., Rechowicz, K., Yamani, Y., and Krusienski, D. J. (2019). Estimating Cognitive Workload in an Interactive Virtual Reality Environment Using EEG. *Front. Hum. Neurosci.* 13. doi:10.3389/fnhum.2019.00401

Tremmel, C., and Krusienski, D. J. (2019). "EEG Spectral Conditioning for Cognitive-State Classification in Interactive Virtual Reality," in 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), 2806–2811.

Tsai, C.-F., Yeh, S.-C., Huang, Y., Wu, Z., Cui, J., and Zheng, L. (2018). The Effect of Augmented Reality and Virtual Reality on Inducing Anxiety for Exposure Therapy: A Comparison Using Heart Rate Variability. *J. Healthc. Eng.* 2018, 6357351. doi:10.1155/2018/6357351

Tsakiris, M., Prabhu, G., and Haggard, P. (2006). Having a Body versus Moving Your Body: How agency Structures Body-Ownership. *Conscious. Cogn.* 15, 423–432. doi:10.1016/j.concog.2005.09.004

Van Gent, P., Farah, H., Nes, N., and van Arem, B. (2018). "Heart Rate Analysis for Human Factors: Development and Validation of an Open Source Toolkit for Noisy Naturalistic Heart Rate Data," in Proceedings of the 6th HUMANIST Conference, 173–178.

Vanderlei, L., Pastre, C., Hoshi, R., Carvalho, T., and Godoy, M. (2009). Basic Notions of Heart Rate Variability and its Clinical Applicability. *Braz. J. Cardiovasc. Surg.* 24, 205–217. doi:10.1590/s0102-76382009000200018

Vince, J. (2004). *Introduction to Virtual Reality*. London, UK: Springer Science & Business Media.

Vinkers, C. H., Penning, R., Hellhammer, J., Verster, J. C., Klaessens, J. H., Olivier, B., et al. (2013). The Effect of Stress on Core and Peripheral Body Temperature in Humans. *Stress* 16, 520–530. doi:10.3109/10253890.2013.807243

Vogt, J., Hagemann, T., and Kastner, M. (2006). The Impact of Workload on Heart Rate and Blood Pressure in En-Route and tower Air Traffic Control. *J. Psychophysiol.* 20, 297–314. doi:10.1027/0269-8803.20.4.297

Volante, M., Babu, S. V., Chaturvedi, H., Newsome, N., Ebrahimi, E., Roy, T., et al. (2016). Effects of Virtual Human Appearance Fidelity on Emotion Contagion in Affective Inter-personal Simulations. *IEEE Trans. Visual. Comput. Graph.* 22, 1326–1335. doi:10.1109/TVCG.2016.2518158

Volonte, M., Hsu, Y., Liu, K., Mazer, J. P., Wong, S., and Babu, S. V. (2020). "Effects of Interacting with a Crowd of Emotional Virtual Humans on Users' Affective and Non-Verbal Behaviors," in 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), 293–302.

Vourvopoulos, A., Pardo, O. M., Lefebvre, S., Neureither, M., Saldana, D., Jahng, E., et al. (2019). Effects of a Brain-Computer Interface With Virtual Reality (VR) Neurofeedback: A Pilot Study in Chronic Stroke Patients. *Front. Hum. Neurosci.* 13. doi:10.3389/fnhum.2019.00210

Wagner, J., Lingenfelser, F., Baur, T., Damian, I., Kistler, F., and André, E. (2013). "The Social Signal Interpretation (Ssi) Framework: Multimodal Signal Processing and Recognition in Real-Time," in Proceedings of the 21st ACM International Conference on Multimedia, 831–834.

Wang, Q., Wang, H., and Hu, F. (2018). "Combining EEG and VR Technology to Assess Fear of Heights," in 2018 9th International Conference on Information Technology in Medicine and Education (ITME), 110–114.

Wang, X., Shi, Y., Zhang, B., and Chiang, Y. (2019). The Influence of Forest Resting Environments on Stress Using Virtual Reality. *Int. J. Environ. Res. Public Health* 16, 3263. doi:10.3390/ijerph16183263

Wang, Y.-G., Liu, M.-H., and Shen, Z.-H. (2019). A Virtual Reality Counterconditioning Procedure to Reduce Methamphetamine Cue-Induced Craving. *J. Psychiatr. Res.* 116, 88–94. doi:10.1016/j.jpsychires.2019.06.007

Wang, Y., Zhai, G., Chen, S., Min, X., Gao, Z., and Song, X. (2019). Assessment of Eye Fatigue Caused by Head-Mounted Displays Using Eye-Tracking. *Biomed. Eng. Online* 18, 111. doi:10.1186/s12938-019-0731-5

Watson, D., Clark, L. A., and Tellegen, A. (1988). Development and Validation of Brief Measures of Positive and Negative Affect: The Panas Scales. *J. Personal. Soc. Psychol.* 54, 1063. doi:10.1037/0022-3514.54.6.1063

Weibel, R. P., Grübel, J., Zhao, H., Thrash, T., Meloni, D., Hölscher, C., et al. (2018). Virtual Reality Experiments with Physiological Measures. *J. Vis. Exp.* (138), 58318. doi:10.3791/58318

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

Wolf, E., Döllinger, N., Mal, D., Wienrich, C., Botsch, M., and Latoschik, M. E. (2020). "Body Weight Perception of Females Using Photorealistic Avatars in Virtual and Augmented Reality," in 2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR).

Wong, C. L., Li, C. K., Chan, C. W. H., Choi, K. C., Chen, J., Yeung, M. T., et al. (2020). Virtual Reality Intervention Targeting Pain and Anxiety Among Pediatric Cancer Patients Undergoing Peripheral Intravenous Cannulation: A Randomized Controlled Trial. *Cancer Nurs.*. [Epub ahead of print]. doi:10.1097/NCC.0000000000000844

Xie, B., Zhang, Y., Huang, H., Ogawa, E., You, T., and Yu, L.-F. (2018). Exercise Intensity-Driven Level Design. *IEEE Trans. Vis. Comp. Graph.* 24, 1661–1670. doi:10.1109/TVCG.2018.2793618

Xu, T., Yin, R., Shu, L., and Xu, X. (2019). "Emotion Recognition Using Frontal EEG in VR Affective Scenes," in 2019 IEEE MTT-S International Microwave Biomedical Conference (IMBioC), 1–4.

Yang, S., He, Y., and Zheng, X. (2019). "FoVR: Attention-Based VR Streaming through Bandwidth-Limited Wireless Networks," in 2019 16th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON), 1–9.

Yeh, S., Li, Y., Zhou, C., Chiu, P., and Chen, J. (2018). Effects of Virtual Reality and Augmented Reality on Induced Anxiety. *IEEE Trans. Neural Syst. Rehabil. Eng.* 26, 1345–1352. doi:10.1109/TNSRE.2018.2844083

Yong-Guang, W., Zhi-Hua, S., and Xuan-Chen, W. (2018). Detection of Patients with Methamphetamine Dependence with Cue-Elicited Heart Rate Variability in a Virtual Social Environment. *Psychiatry Res.* 270, 382–388. doi:10.1016/j.psychres.2018.10.009

Yoo, S., Parker, C., and Kay, J. (2018). "Adapting Data from Physical Activity Sensors for Visualising Exertion in Virtual Reality Games," in Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers UbiComp '18, Singapore, Singapore (New York, NY: Association for Computing Machinery), 307–310.

Yoshimura, A., Khokhar, A., and Borst, C. W. (2019). "Eye-Gaze-Triggered Visual Cues to Restore Attention in Educational VR," in 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), 1255–1256.

Yu, C.-P., Lee, H.-Y., and Luo, X.-Y. (2018). The Effect of Virtual Reality forest and Urban Environments on Physiological and Psychological Responses. *Urban For. Urban Green.* 35, 106–114. doi:10.1016/j.ufug.2018.08.013

Zeng, N., Pope, Z., and Gao, Z. (2017). Acute Effect of Virtual Reality Exercise Bike Games on College Students' Physiological and Psychological Outcomes. *Cyberpsychology Behav. Soc. Netw.* 20, 453–457. doi:10.1089/cyber.2017.0042

Zhang, S., Zhang, Y., Sun, Y., Thakor, N., and Bezerianos, A. (2017). "Graph Theoretical Analysis of EEG Functional Network during Multi-Workload Flight Simulation experiment in Virtual Reality Environment," in 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 3957–3960.

Zhang, W., Shu, L., Xu, X., and Liao, D. (2017). "Affective Virtual Reality System (Avrs): Design and Ratings of Affective Vr Scenes," in 2017 International Conference on Virtual Reality and Visualization (ICVRV) (IEEE), 311–314.

Zheng, L. J., Mountstephens, J., and Teo, J. (2020). Four-Class Emotion Classification in Virtual Reality Using Pupillometry. *J. Big Data* 7, 1–9. doi:10.1186/s40537-020-00322-9

Zhu, L., Tian, X., Xu, X., and Shu, L. (2019). "Design and Evaluation of the Mental Relaxation VR Scenes Using Forehead EEG Features," in 2019 IEEE MTT-S International Microwave Biomedical Conference (IMBioC), 1–4.

Zimmer, P., Buttlar, B., Halbeisen, G., Walther, E., and Domes, G. (2019). Virtually Stressed? A Refined Virtual Reality Adaptation of the Trier Social Stress Test (TSST) Induces Robust Endocrine Responses. *Psychoneuroendocrinology* 101, 186–192. doi:10.1016/j.psyneuen.2018.11.010

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Virtual Reality

PERSPECTIVE
published: 15 July 2021
doi: 10.3389/frvir.2021.693327

Perspective: Does Realism Improve Presence in VR? Suggesting a Model and Metric for VR Experience Evaluation

Sungchul Jung and Robert. W Lindeman*

Human Interface Technology Lab (HIT Lab NZ), College of Engineering, University of Canterbury, Christchurch, New Zealand

The concepts of "immersion" and "presence" have been considered as staple metrics for evaluating the quality of virtual reality experiences for more than five decades, even as the concepts themselves have evolved in terms of both technical and psychological aspects. To enhance the user's experience, studies have investigated the impact of different visual, auditory, and haptic stimuli in various contexts to mainly explore the concepts of "plausibility illusion" and "place illusion". Previous research has sometimes shown a positive correlation between increased realism and an increase in presence, but not always, and thus, very little of the work around the topic of presence reports an unequivocal correlation. Indeed, one might classify the overall findings within the field around presence as "messy". Better (or more) visual, auditory, or haptic cues, or increased agency, may lead to increased realism, but not necessarily increased presence, and may well depend on the application context. Rich visual and audio cues in concert contribute significantly to both realism and presence, but the addition of tactile cues, gesture input support, or a combination of these might improve realism, but not necessarily presence. In this paper, we review previous research and suggest a possible theory to better define the relationship between increases in sensory-based realism and presence, and thus help VR researchers create more effective experiences.

Keywords: realism, presence, immersion, evaluation model, metric, illusion, theory, Coherence

OPEN ACCESS

Edited by:
Amela Sadagic,
Naval Postgraduate School,
United States

Reviewed by:
Richard Skarbez,
La Trobe University, Australia
Aryabrata Basu,
Emory University, United States

****Correspondence:***
Sungchul Jung
sungchul.jung@canterbury.ac.nz

Specialty section:
This article was submitted to Virtual Reality and Human Behaviour, a section of the journal Frontiers in Virtual Reality

Received: *10 April 2021*
Accepted: *30 June 2021*
Published: *15 July 2021*

Citation:
Jung S and Lindeman RW (2021) Perspective: Does Realism Improve Presence in VR? Suggesting a Model and Metric for VR Experience Evaluation. Front. Virtual Real. 2:693327. doi: 10.3389/frvir.2021.693327

1 INTRODUCTION

"It's so real!" "This is such a realistic experience!" We believe almost all Virtual Reality (VR) researchers and developers have heard these expressions at least once when successfully delivering immersive VR experience to the general public. Similarly, responses such as "It seems like I'm in another place", or "This looks like my body" are often expressed by the users. General users might not care about the academic distinctions between immersion and illusion, and might even seem confused by them. But clearly, these kinds of responses both implicitly or explicitly indicate the general quality of the VR experience, in a positive direction. Under current circumstances, the most frequently used term is probably related to realism (or realness), and thus using the phrase "the level of realism" might be the easiest way to make the general public understand the quality of a VR experience. On the other hand, in academia, more precise terms, such as immersion (tele-, co-, etc.), presence, embodiment, and body-ownership have been suggested to more-precisely define and comprehend how to evaluate the quality of VR experiences (Meehan et al., 2002; Bowman and McMahan, 2007; Slater, 2009; Kilteni et al., 2015; Skarbez et al., 2017).

The two most common high-level concepts, immersion, and place illusion, have been studied along defined lines of objective and subjective aspects, respectively. Immersion is often defined as an objective property Slater (1999), Bowman and McMahan (2007) of a VR system's profile. For example, the visual stimuli from the screen size, resolution, stereo, field-of-view, head-tracked head-mounted display (HMD) with full real-time motion capture are critical to simulate a computer-generated experience to the user. Researchers concluded that a system that provides rich virtual surroundings, along with the user's own body movements such as looking around or reaching out to touch a certain object, provides a higher level of immersion than a system that does not support such visual dynamics along with the user's movement. Thus, a VR experience that uses a HMD has higher immersion than a screen-based VR experience. In line with this perspective, *realness* can be evaluated by, for example, the visual representation of the computer-generated world in terms of the number of triangles and the resolution of textures McDonnell et al. (2012), Latoschik et al. (2017), and enhanced auditory feedback such as spatialized audio (Naef et al., 2002). In addition, supporting the expected additional sensory channels in line with the context of the given VR experience, such as tactile, olfactory, and taste feedback, can increase the sense of realness (Feng et al., 2015; Feng et al., 2016; Jung et al., 2020; Jung et al., 2021a; Jung et al., 2021b).

In contrast, *illusion* is related to how humans subjectively perceive an immersive experiences: illusion is regarded in terms of psychological aspects (Slater, 2009). To evaluate the quality of these subjective aspects of VR, researchers developed the concepts of presence and body ownership based on the perception or cognition of objects, whether regarding the surroundings or their own avatar body representation, using questionnaire-based constructs or physiological signals (Slater, 1999; Meehan et al., 2002; Slater et al., 2010; Latoschik et al., 2017). The sense of body ownership is regarded as a clear construct, and so at a high level, it may influence the quality of VR experiences. Thus, researchers generally see a positive relationship between it and presence. However, there is still a vague gap in the nature of this relationship, related to the various definitions and understandings of presence. To help address this confusion, two subsequent concepts, *place* illusion and plausibility illusion, have been suggested (Slater, 2009; Skarbez, 2016). Along with these methods for evaluating the subjective quality of VR experiences, deciding how different levels of immersion might correspond to different levels of illusion has been a focus, measuring how much subjects' responses to events in the virtual world matched reactions to those in the real world (Slater, 2009; Skarbez et al., 2017). In other words, the perception or cognition of the surroundings and their own virtual body are correlated, and thus highly influence the sense of presence and body ownership, leading to subjectively highly-rated qualified VR experiences.

Using these terms and definitions as lenses, revisiting the two opening user comments, "It is so real!" and "This is such a realistic experience!", might be interesting. How can we interpret them? Should we handle them as part of immersion or as illusion comments? Does improved realism also enhance the sense of *illusion?* In the light of our perspective, a deeper understanding of the term "Realness" or "Realism" is required, since those terms can be situated in both the immersion or illusion aspects, depending on the definition. Realism could be handled as an immersion component if we define it as the extent and quality of the sensory channels. For example, multisensory stimuli improve VR systems; that is, they help provide additional sensory information, such as tactile, olfactory, or taste, which should theoretically lead to a more-immersive VR system compared to traditional VR systems that typically provide only visual and audio cues. In 1999, based on earlier work by Hinckley et al. (1994), Lindeman et al. (1999) introduced the concept of Passive Haptics, mainly related to the use of hand-held props, and observed enhanced performance for manipulating interface tools in virtual environments. Similarly, Insko (2001) designed a study to measure the sense of presence when passive haptics were used. Participants were asked to stand on a 5 cm physical ledge while they stood on a 20 m pit in a VR room. The physical ledge allowed users to leverage the feeling of passive haptics, and the researchers found a significant increase in sense of presence. In addition to visual and audio feedback, the use of highly-congruent haptic feedback showed positive impact on the sense of presence.

In 1999, Dinh et al. (1999) investigated the impact of multisensory VR experiences, using a large number of sensory modalities including tactile, olfactory, audio, and visual cues. They found that increasing the number of modalities of sensory input in a virtual environment can increase the sense of presence. However, they reported that increased visual realism did not. After 2 decades, Jung et al. designed a system to deliver additional sensory feedback (vibration, wind, and olfactory cues) in multiple studies (Jung et al., 2020; Jung et al., 2021a; Jung et al., 2021b). In each of these cases, systems with multisensory cues can be seen as (objectively) more immersive. However, if we consider realism within some contextual fidelity, it could be interpreted as an illusionary component.

In this short article, we propose a research question and then present our perspective on this question based on previous research and our own experiences.

Note: This article is based on the assumption that VR has at least visual stimuli, regardless existence of other external sensory stimulation (Skarbez et al., 2021), following the definition of Milgram and Kishino's reality-virtuality continuum (Milgram and Kishino, 1994).

2 LITERATURE REVIEW FOR IMMERSION AND ILLUSION

The terms presence and immersion have been suggested and used actively in the VR community or even by the general public sometimes, as representative of the quality of a VR experience, or of the expected outcome of exposure to a VR experience. Depending on the technical and psychological context, as well as aspects of the VR experience, many interrelated terms have been used, such as presence, co-presence, tele-presence, social presence, embodiment, and body ownership. Our goal in this section is to revisit the most important terms, while introducing two other terms, Coherence and Realism, that have not been studied as deeply, but that we feel are important. Finally, we suggest a possible model to show the correlation between them in the following next section.

2.1 Immersion

Witmer and Singer state that immersion as a subjectively perceived psychological characteristic to the surrounding environment and events in the computer-generated world (Witmer and Singer, 1998). However, and Bowman and McMahan argue that immersion is a systemically objective characteristic of a VR system (Slater, 2009; Bowman and McMahan, 2007). Similarly, Lombard et al. categorized immersion into perceptual immersion and psychological immersion (Lombard et al., 2000). Due to these seemingly-contradictory perspectives of this single term, a clear single point of reference is definitely required in the community. In this article, we support Slater's perspective, that immersion refers to an objective characteristic of a VR system, and address illusion in 2.3.

Our definition of immersion is objective and has clear metrics: more is better. Wide and high-resolution field-of-view (FOV) HMDs provide more immersion than screen-based VR. Also, wide and accurate six-degree-of-freedom (DOF) tracking provides more immersion than three DOF tracking. Similarly, spatial audio cues are more immersion than binaural or monaural audio cues. In light of this perspective, adding more sensory channels should, theoretically, raise the level of immersion. For example, extending secondary feedback cues, such as floor or wearable vibration, wind, and olfactory stimuli that match the visual and audio stimuli, provides more immersion than a system with visual and audio only (Feng et al., 2015; Feng et al., 2016; Jung et al., 2020; Jung et al., 2021a; Jung et al., 2021b).

2.2 Coherence

Skarbez et al. first defined the term Coherence to mean the set of reasonable circumstances that can lead to a convincing context without additional explanation, based on a Bayesian prior (Samad et al., 2015; Skarbez, 2016). Coherence is the quality of internal logic and behavioral consistency of a VR experience; thus, the notion does not depend on the faithful representation of real-world experiences (Skarbez et al., 2017). For example, supernatural abilities such as teleportation, invisibility, or flying experienced in the context of science-fiction or fantasy would be regarded as coherent behavior.

2.2.1 Realism

Realism or fidelity can be described in as the extent to which the virtual environment emulates the real world (Alexander et al., 2005). Because of the interchangeable usage of the two terms, we use Realism in this article. Similarly, Stoffregen et al. observed that highly realistic systems can produce sensory stimuli that are identical to real-world stimuli, and thus stimulus fidelity can be regarded as an objective characteristic of a simulation. Riccio et al. described experiential fidelity as a subjective experience while *action fidelity* is a systemic performance (Riccio, 1995). Based on these definitions, conceptualized Coherence as a superset of Realism (Skarbez et al., 2017). Thus, considering the given aspects of the correlation between Realism and Coherence, it is still an open question as to whether Coherence can be regarded as an objective construct or not. As a possible solution for handling Coherence as an objective measure, Skarbez et al. (2017) redefined the domain of the construct "as the set of objectively reasonable circumstances that can be demonstrated by the scenario without introducing objectively unreasonable circumstances." This redefinition is aligned with the previously suggested notion of *Experiential Fidelity* proposed by Beckhaus and Lindeman, 2011).

2.2.2 Constructs of Coherence

Alexander et al. suggest that Realism (or Fidelity) has three subcategories: physical simulation, functional simulation, and psychological fidelity (Alexander et al., 2005). *Physical simulation* refers to the operational environment such as fully multisensory-enabled experience including visual, aural, tactile, olfaction, and taste stimuli. For example, (Feng et al., 2015; Feng et al., 2016) and (Jung et al., 2020; Jung et al., 2021a; Jung et al., 2021b) achieved a highly-realistic VR experience by using multisensory-cue-enabled VR platforms. *Functional simulation* is the fidelity of the behavioral representation by the operational equipment in reacting to the tasks executed by the user. While these two components are related to the stimuli from the system, the third one, *psychological fidelity*, is related to the faithfulness of the psychological effects that the simulation creates with regard to those that would be experienced in a real-world version of the experience.

2.3 Illusion

Based on our adoption of Slater's definition of immersion, we categorize Witmer and Singer's description of *Illusion* as the subjective perception of the psychological characteristic of the surrounding environment and events in the VR environment. In this article, we use *Illusion* as an equivalent term for *Presence* that can represent the overall quality of the virtual experience subjectively. For example, Presence is most commonly defined as the feeling of "being there" in a virtual place (Witmer and Singer, 1998). While Presence refers to the feeling of the surrounding environment and events to the first-person's egocentric experience, this notion can be extended to the perception or cognition of other entities' existence, and we call this *Co-presence*. The concept generally is defined as "being there together" (Schroeder, 2005). Specifically, Goffman et al. stated that Co-presence is related to the sense of mutual perception between two or more (Goffman, 1963). *Social-presence* also has been suggested to mean the feeling of awareness of being present with other entities with the degree of attention level (Nowak and Biocca, 2003). In this article, we follow and define Social-presence as the moment-by-moment awareness of the Co-presence of another sentient being accompanied by a sense of engagement with them Skarbez et al. (2017), and thus we would like to coin a simple representation, "Being there, engaged together," to encapsulate social-presence.

2.3.1 Constructs of Illusion

As constructs of Illusion, two logically orthogonal aspects, Place Illusion (PI) and Plausibility Illusion (Psi), have been suggested by (Slater, 2009). He defines PI as "the illusion of being in a place in spite of the sure knowledge that you are not there," and Psi as "the illusion that what is apparently happening is really happening even though you know for sure that it is not". In short, according to this definition, PI can be mapped to the conventional notion of the spatial Presence, while the Psi indicates ones belief that the events happening are things one is actually experiencing (Skarbez et al., 2017).

Embodiment, a sense of "having a body", has been suggested to be a mostly subjective feeling in psychology (Kilteni et al., 2015). Based on the definition of embodiment, virtual body ownership, the sense of feeling "ownership of the given virtual body (of part or whole)" has been developed and researched through numerous studies Banakou et al. (2013), and is regarded as a critical metric for VR experiences, with an implicit agreement on the existence of a correlation with Presence (Yuan and Steed, 2010). To enhance the sense of body ownership in VR, visuo-motor, visuo-tactile, anatomical plausibility Kilteni et al. (2015), and personalized avatar appearance, regardless of whether given directly or indirectly, have been suggested as critical components (Jung and Hughes, 2016; Jung et al., 2017; Jung et al., 2018; Waltemate et al., 2018).

3 POSSIBLE CORRELATION MODEL FOR IMMERSION AND ILLUSION

Following Slater's approach for evaluating the quality of VR experience seems to suggest a clear differentiation between Immersion and Illusion (or Presence). However, researchers still observe and report some non-orthogonality between these two high-level concepts. This might be caused due by 1) ambiguity of sub-component definitions, and 2) a possibility of the existence of other factors. For example, *Realism* can be interpreted either as influenced by either immersion or illusion, depending on the VR context. While the term Realism has been used by both the general public and academia, a clear definition of Realism in VR has not been given. Alternatively, *Fidelity* is regarded as a similar concept, and Alexander et al. describes Fidelity as the extent to which the virtual environment emulates the real world Alexander et al. (2005) in terms of functional, physical, and psychological perspectives. On the other hand, Stoffregen et al. focus on the sensory stimuli provided by the system (Stoffregen et al., 2003). We believe that Alexander's definition includes Stoffregen's definition, and so prefer Alexander's viewpoint regarding Realism as have done (Skarbez et al., 2017).

Meanwhile, suggest a revised version of a reality continuum, and propose a three-dimensional model that includes "Extent of World Knowledge," "Immersion," and "Coherence." A deep discussion of this important work is beyond the scope of this paper; please refer to it for more details (Skarbez et al., 2021). Noticeable attributes of this model are that Coherence is similar to Psi, and both suggest the importance of the context of the given VR experience. However, Coherence might not be limited to a psychological context. Following Skarbez's description of the term Coherence (Skarbez et al., 2017), we accept that Realism could be a subset of Coherence. If we recall the definition of Realism (or Fidelity) by Alexander et al., it can be said to be comprised of three components: Physical Fidelity, Functional Fidelity, and Psychological Fidelity. Thus, logically, Coherence involves those aspects too.

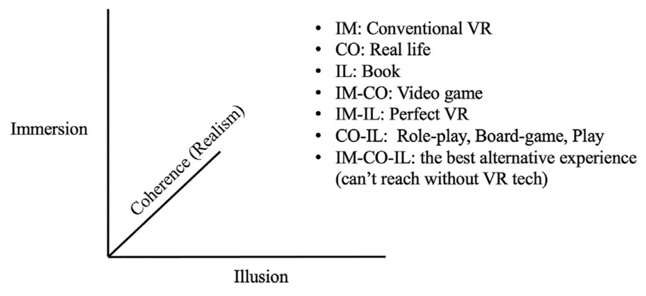

FIGURE 1 | Model for Quality of VR Experience using Orthogonal Dimensions of Coherence (CO), Immersion (IM), and Illusion (IL).

Based on these definitions, we suggest Coherence (which encapsulates Realism), Immersion and Illusion as forming orthogonal axes, creating a comprehensive VR experience evaluation model (**Figure 1**). We can use this to describe the subjective feelings that arise in a user who encounters experiences placed along each axis within this space, including when supports are maximally provided. However, even though the given dimensions are orthogonal, and thus should not interact with each other, it is a challenge to measure the feeling even if we successfully provide a controlled experience. From a practical perspective, the orthogonal model could be represented using a Venn diagram approach, as can be seen in **Figure 2**. In this model, we also provide suggested or validated constructs for measuring the lower-level components. Based on the suggested Venn model, we argue that increased Realism can improve the chances of achieving feelings of deep Presence, including Co-presence and Social Presence, partially if the given Realism satisfied context, but not necessarily. Of course, measuring method should be considered but it is beyond our scope in this article.

3.1 Implications

Considering the proposed orthogonal model, Coherence that incorporates Realism does not directly enhance Illusion and Immersion. Improving sensory realism incurs a high cost due to the required hardware support. In this case, what is the motivation for providing realistic systems, such as multisensory-enabled platforms?

In our own work, we have repeatedly reported higher *preference* responses for our multisensory VR systems compared to typical VR systems, regardless of the context, number of sensory channels, and level of fidelity, even though multisensory cues did not consistently

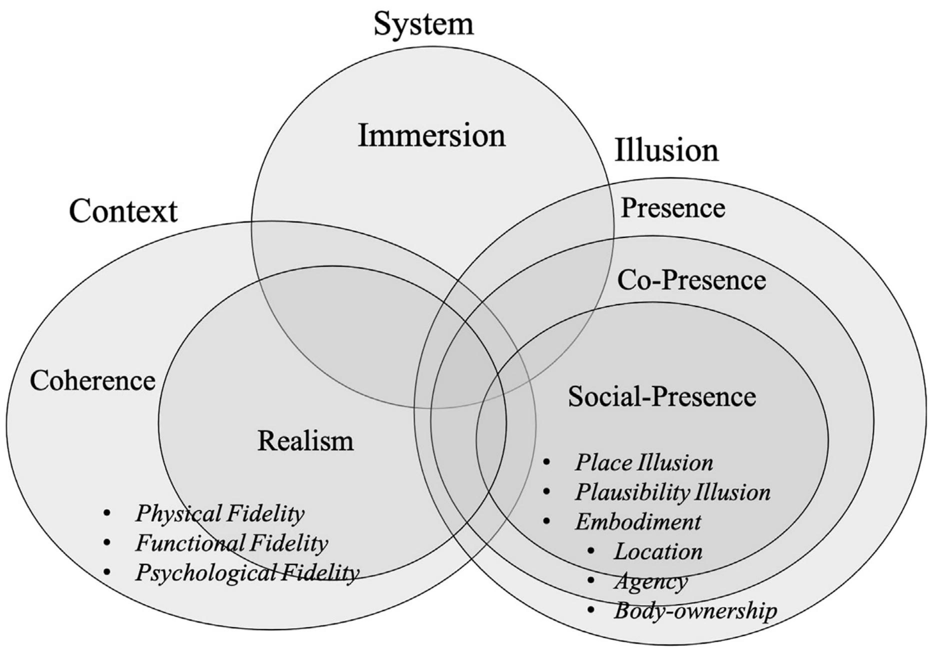

FIGURE 2 | Venn diagram model for quality of VR experience.

lead to higher Presence (Jung et al., 2020; Jung et al., 2021b; Jung et al., 2021a). The reasons are still not clear to us, but participants reported they felt a stronger sense of *Fun*, *Impressiveness* and *Involvement* in multisensory VR, which are strongly related to emotional engagement in the given place, object, and maybe certain events too, which sounds a lot like *Aura* (MacIntyre et al., 2004). Similarly, Doukakis et al. (2019) reported that visual feedback was the dominant factor chosen for designing VR on a limited budget. However, as the budget size increases, the preference for having a balanced distribution of resources (e.g., having additional smell feedback) increased. Thus, based on the observed and reported trends, we claim VR experiences with more and higher-quality stimuli fed to sensory systems might have stronger preference, and thus we suggest *Preference* as a new metric to evaluate VR experience as an exclusive factor from the proposed model. This is because the suggested model accounts for experiential attributes, but not other factors, such as cost, encumbrance, fun, or engagement. It is possible, for example, to have photorealistic graphics in gaming experiences that do not guarantee fun. The Nintendo Switch has a low-end system profile compared to the PlayStation or Xbox series in general, but user preference is similar among the systems. Thus we conclude that the proposed axis model works as a specific tool to evaluate the VR experience, along with three independent axes, and so we can evaluate solo VR experiences as well. On the other hand, Preference is a comparison tool that depends on the user's choice which might come from the overall experience compared to other given VR experiences. Most research does not tend to empathize the importance of *User Preference* in studies on VR experience. However, we argue that Preference can be a critical indicator in terms of business perspectives, since it clearly shows the overall evaluation from a direct comparison, and thus connects to people's preferences for future VR experiences.

4 CONCLUSION

In this article, we have reflected on two key high-level concepts, Immersion and Illusion. Based on previous work, we revisited the concepts of Coherence and Realism, and how they correlate with Illusion (as a representative of the sense of Presence) and Immersion. In order to explore these questions, we refer to a series of studies around the influence of multisensory cues in VR, and finally proposed two new models for representing the relationship between these three components, Coherence, Immersion, and Illusion. We conclude that these can be treated as independent dimensions, but that they might partially influence each other, as they intersect. Finally, we also suggest that researchers and designers consider Preference as a critical component for evaluating the impact of VR experiences, especially from a business perspective.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTION

SJ and RL contributed to the ideation and development of this article. SJ developed the initial draft and RL contributed to further revision and polishing of the final article.

REFERENCES

Alexander, A., Brunyé, T. T., Sidman, J., and Weil, S. A. (2005). *From Gaming to Training: A Review of Studies on Fidelity, Immersion, Presence, and Buy-In and Their Effects on Transfer in Pc-Based Simulations and Games.*

Banakou, D., Groten, R., and Slater, M. (2013). Illusory Ownership of a Virtual Child Body Causes Overestimation of Object Sizes and Implicit Attitude Changes. *Proc. Natl. Acad. Sci.* 110, 12846–12851. doi:10.1073/pnas.1306779110

Beckhaus, S., and Lindeman, R. W. (2011). *Experiential Fidelity: Leveraging the Mind to Improve the VR Experience*. Vienna: Springer Vienna, 39–49. doi:10.1007/978-3-211-99178-7_3

Bowman, D. A., and McMahan, R. P. (2007). Virtual Reality: How Much Immersion Is Enough? *Computer* 40, 36–43. doi:10.1109/MC.2007.257

Dinh, H. Q., Walker, N., Hodges, L. F., Chang Song, C., and Kobayashi, A. (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), Houston, TX, USA, 13-17 March 1999, 222–228. doi:10.1109/VR.1999.756955

Doukakis, E., Debattista, K., Bashford-Rogers, T., Dhokia, A., Asadipour, A., Chalmers, A., et al. (2019). Audio-visual-olfactory Resource Allocation for Tri-modal Virtual Environments. *IEEE Trans. Visualization Comput. Graphics* 25, 1865–1875. doi:10.1109/tvcg.2019.2898823

Feng, M., Dey, A., and Lindeman, R. W. (2016). "An Initial Exploration of a Multi-Sensory Design Space: Tactile Support for Walking in Immersive Virtual Environments," in 2016 IEEE Symposium on 3D User Interfaces (3DUI), Greenville, SC, USA, 19-20 March 2016, 95–104. doi:10.1109/3DUI.2016.7460037

Feng, M., Lindeman, R. W., Abdel-Moati, H., and Lindeman, J. C. (2015). "Haptic ChairIO: A System to Study the Effect of Wind and Floor Vibration Feedback on Spatial Orientation in VEs." In 2015 IEEE Symposium on 3D User Interfaces (3DUI), Arles, France, 23-24 March 2015, 149–150. doi:10.1109/3DUI.2015.7131744

Goffman, E. (1963). *Behavior in Public Places ; Notes on the Social Organization of Gatherings / Erving Goffman*. New York: Free Press.

Hinckley, K., Pausch, R., Goble, J. C., and Kassell, N. F. (1994). "Passive Real-World Interface Props for Neurosurgical Visualization," in CHI '94 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Boston, MA (New York, NY, USA: Association for Computing Machinery), 452–458. doi:10.1145/191666.191821

Insko, B. E. (2001). Passive Haptics Significantly Enhances Virtual Environments. Ph.D. thesis (Chapel Hill, NC: The University of North Carolina at Chapel Hill). AAI3007820.

Jung, S., and Hughes, C. E. (2016). "The Effects of Indirect Real Body Cues of Irrelevant Parts on Virtual Body Ownership and Presence," in ICAT-EGVE '16 Proceedings of the 26th International Conference on Artificial Reality and Telexistence and the 21st Eurographics Symposium on Virtual Environments, Little Rock, AR (Goslar, DEU: Eurographics Association), 107–114.

Jung, S., Karki, N., Slutter, M., and Lindeman, R. W. (2021a). On the Use of Multi-Sensory Cues in Symmetric and Asymmetric Shared Collaborative Virtual Spaces. *Proc. ACM Hum.-Comput. Interact* 5, 1-25. doi:10.1145/3449146

Jung, S., Li, R., McKee, R., Whitton, M. C., and Lindeman, R. W. (2021b). "Floor-vibration Vr: Mitigating Cybersickness Using Whole-Body Tactile Stimuli in Highly Realistic Vehicle Driving Experiences," in 2021 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Lisbon, Portugal.

Jung, S., Sandor, C., Wisniewski, P. J., and Hughes, C. E. (2017). "Realme: The Influence of Body and Hand Representations on Body Ownership and Presence," in SUI '17 Proceedings of the 5th Symposium on Spatial User Interaction, Brighton, United Kingdom (New York, NY, USA: Association for Computing Machinery), 3–11. doi:10.1145/3131277.3132186

Jung, S., Wisniewski, P. J., and Hughes, C. E. (2018). "In Limbo: The Effect of Gradual Visual Transition between Real and Virtual on Virtual Body Ownership Illusion and Presence," in 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Tuebingen/Reutlingen, Germany, 18-22 March 2018, 267–272. doi:10.1109/VR.2018.8447562

Jung, S., Wood, A. L., Hoermann, S., Abhayawardhana, P. L., and Lindeman, R. W. (2020). "The Impact of Multi-Sensory Stimuli on Confidence Levels for Perceptual-Cognitive Tasks in Vr," in 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Atlanta, GA, USA, 22-26 March 2020, 463–472. doi:10.1109/VR46266.2020.00067

Kilteni, K., Maselli, A., Kording, K. P., and Slater, M. (2015). Over My Fake Body: Body Ownership Illusions for Studying the Multisensory Basis of Own-Body Perception. *Front. Hum. Neurosci.* 9, 141. doi:10.3389/fnhum.2015.00141

Latoschik, M. E., Roth, D., Gall, D., Achenbach, J., Waltemate, T., and Botsch, M. (2017). "The Effect of Avatar Realism in Immersive Social Virtual Realities," in Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology.

Lindeman, R., Sibert, J., and Hahn, J. (1999). "Hand-held Windows: towards Effective 2d Interaction in Immersive Virtual Environments," in Proceedings IEEE Virtual Reality (Cat. No. 99CB36316), Houston, TX, USA, 13-17 March 1999, 205–212. doi:10.1109/VR.1999.756952

Lombard, M., Ditton, T., Crane, D., Davis, B., Gil-Egui, G., Horvath, K., et al. (2000). "Measuring Presence: A Literature-Based Approach to the Development of a Standardized Paper-And-Pencil Instrument," in Third International Workshop on Presence, Delft, Netherlands.

MacIntyre, B., Bolter, J., and Gandy, M. (2004). "Presence and the Aura of Meaningful Places," in 7th Annual International Workshop on Presence (PRESENCE 2004). Polytechnic University of Valencia: Valencia, Spain.

McDonnell, R., Breidt, M., and Bülthoff, H. H. (2012). Render Me Real? Investigating the Effect of Render Style on the Perception of Animated Virtual Humans. *ACM Trans. Graph.* 31, 1-11. doi:10.1145/2185520.2185587

Meehan, M., Insko, B., Whitton, M., and Brooks, F. P. (2002). Physiological Measures of Presence in Stressful Virtual Environments. *ACM Trans. Graph.* 21, 645–652. doi:10.1145/566654.566630

Milgram, P., and Kishino, F. (1994). A Taxonomy of Mixed Reality Visual Displays. *IEICE Trans. Inf. Syst.* 77, 1321–1329.

Naef, M., Staadt, O., and Gross, M. (2002). "Spatialized Audio Rendering for Immersive Virtual Environments," in VRST '02 Proceedings of the ACM Symposium on Virtual Reality Software and Technology, Hong Kong, China (New York, NY, USA: Association for Computing Machinery), 65–72. doi:10.1145/585740.585752

Nowak, K. L., and Biocca, F. (2003). The Effect of the agency and Anthropomorphism of Users' Sense of Telepresence, Copresence, and Social Presence in Virtual Environments. *Presence: Teleoper. Virtual Environ.* 12, 481–494. doi:10.1162/105474603322761289

Riccio, G. E. (1995). "Coordination of Postural Control and Vehicular Control: Implications for Multimodal Perception and Simulation of Self-Motion," in *Local Applications of the Ecological Approach to Human-Machine Systems*. Editor J. C. K. V. Peter Hancock (John Flach), 60.

Samad, M., Chung, A. J., and Shams, L. (2015). Perception of Body Ownership Is Driven by Bayesian Sensory Inference. *PLOS ONE* 10, 1–23. doi:10.1371/journal.pone.0117178

Schroeder, R. (2005). "Copresence and Interaction in Virtual Environments: An Overview of the Range of Issues," in Presence 2002: Fifth International Workshop, Porto, Portugal, 274–295.

Skarbez, R., Brooks, F. P., Jr., and Whitton, M. C. (2017). A Survey of Presence and Related Concepts. *ACM Comput. Surv.* 50, 1-39. doi:10.1145/3134301

Skarbez, R. (2016). *Plausibility Illusion in Virtual Environments*. Chapel Hill, NC: University of North Carolina at Chapel Hill. doi:10.17615/2mg3-gh93

Skarbez, R., Smith, M., and Whitton, M. C. (2021). Revisiting Milgram and Kishino's Reality-Virtuality Continuum. *Front. Virtual Reality* 2, 27. doi:10.3389/frvir.2a021.64799710.3389/frvir.2021.647997

Slater, M. (1999). Measuring Presence: A Response to the Witmer and Singer Presence Questionnaire. *Presence: Teleoper. Virtual Environ.* 8, 560–565. doi:10.1162/105474699566477

Slater, M. (2009). Place Illusion and Plausibility Can lead to Realistic Behaviour in Immersive Virtual Environments. *Philos. Trans. R. Soc. B: Biol. Sci.* 364, 3549–3557. doi:10.1098/rstb.2009.0138

Slater, M., Spanlang, B., and Corominas, D. (2010). Simulating Virtual Environments within Virtual Environments as the Basis for a Psychophysics of Presence. *ACM Trans. Graph.* 29, 1-9. doi:10.1145/1778765.1778829

Stoffregen, T. A., Bardy, B. G., Smart, L. J., and Pagulayan, R. J. (2003). "On the Nature and Evaluation of Fidelity in Virtual Environments: Applications, Implications, and Human Performance," in *Virtual and Adaptive Environments*. Editors L. J. Hettinger and M. W. Haas (Mahwah, NJ: Lawrence Erlbaum Associates, Inc), 111–127.

Waltemate, T., Gall, D., Roth, D., Botsch, M., and Latoschik, M. E. (2018). The Impact of Avatar Personalization and Immersion on Virtual Body Ownership, Presence, and Emotional Response. *IEEE Trans. Visualization Comput. Graphics* 24, 1643–1652. doi:10.1109/TVCG.2018.2794629

Witmer, B. G., and Singer, M. J. (1998). Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments* 7, 225–240. doi:10.1162/105474698565686

Yuan, Y., and Steed, A. (2010). "Is the Rubber Hand Illusion Induced by Immersive Virtual Reality?," in 2010 IEEE Virtual Reality Conference (VR), Boston, MA, USA, 20-24 March 2010, 95–102. doi:10.1109/VR.2010.5444807

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Virtual Reality

ORIGINAL RESEARCH
published: 12 August 2021
doi: 10.3389/frvir.2021.636643

Reality Stems From Modality: Stereotype Threat Effects of a STEM Game in Augmented and Virtual Reality

Rabindra Ratan[1]*, *Josephine K. Boumis*[2], *Sarah Kuang*[3], *Andrew Gambino*[4] and *Kuo-Ting Huang*[5]

[1]Department of Media and Information, Michigan State University, East Lansing, MI, United States, [2]Department of Communication, University of South Florida, Tampa, FL, United States, [3]Department of Psychology, Michigan State University, East Lansing, MI, United States, [4]College of Communications, Pennsylvania State University, State College, PA, United States, [5]Department of Journalism, Ball State University, Muncie, IN, United States

This study examined the relationship between stereotype threat, game modality (augmented reality, virtual reality), and stereotypic beliefs about STEM fields. Results of a 2 [modality] x 2 [stereotype threat] factorial, between-subjects experiment with women participants (N = 64) suggest that gender stereotypes primed before playing the STEM game in AR induced stereotype threat, but induced stereotype reactance in VR. Specifically, for participants who played in AR, the stereotype-reinforcing prompt (compared to a counter-stereotype prompt) was associated with worse STEM-game performance, which mediated an increase in stereotypical beliefs about women in STEM. Conversely, for participants who played in VR, the stereotype-reinforcing prompt was associated with better STEM-game performance and more positive (i.e., counter-stereotypic) beliefs about women in STEM, though without mediation. These findings support the claim that stereotypes triggered in a STEM-gaming context have the potential to reinforce stereotypes in STEM fields. Researchers and practitioners should consider the implication that VR is potentially more male-stereotyped than AR, while AR makes stereotyped identity characteristics more accessible than VR.

Keywords: augmented and virtual reality, experiment, video games, STEM games, stereotype threat and reactance

OPEN ACCESS

Edited by:
Richard Skarbez,
La Trobe University, Australia

Reviewed by:
Kelly Lynn Mulvey,
North Carolina State University, United States
Harry Farmer,
University of Greenwich, United Kingdom

Correspondence:
Rabindra Ratan
rar@msu.edu

Specialty section:
This article was submitted to Virtual Reality and Human Behavior, a section of the journal Frontiers in Virtual Reality

Received: *01 December 2020*
Accepted: *03 May 2021*
Published: *12 August 2021*

Citation:
Ratan R, Boumis JK, Kuang S, Gambino A and Huang K-T (2021) Reality Stems From Modality: Stereotype Threat Effects of a STEM Game in Augmented and Virtual Reality. Front. Virtual Real. 2:636643. doi: 10.3389/frvir.2021.636643

INTRODUCTION

Although the percentage of women in science, technology, engineering, math (STEM) fields has been increasing worldwide (Wiest et al., 2017), gender inequality in STEM fields is still a major problem in the U.S., with women only representing 24% of the STEM workforce (Noonan, 2017). Organizations and programs that encourage more students, especially girls, toward STEM careers often utilize activities related to video games (Jenson et al., 2007; Collette, 2013), which act as an entry point to STEM thinking (e.g., design, programming; Giammarco et al., 2015). The present research extends an understanding of the relationship between video games and STEM to the understudied context of augmented reality (AR) and virtual reality (VR) gaming. Given their immersive nature, AR and VR are potentially better at facilitating STEM-relevant skills, such as spatial rotation ability (Spence and Feng, 2010; Granic et al., 2014), compared to traditional (e.g., flatscreen) gaming modalities. Although AR and VR both provide immersive experiences that can enhance learning outcomes, they are also fundamentally different in the ways that they present and allow users to interact with educational information, which may influence learning outcomes. Building from a recent finding that AR leads to better retention of auditory content, while VR is better for visual content (Huang et al., 2019), the present research compares VR and AR modalities with a common facet of video games and STEM fields: gender[1] stereotypes.

Just as STEM fields are male-dominated, male-catering, and often hostile environments for women (Smith et al., 2013), women also receive more negative commentary in gaming contexts than men, regardless of skill (Kuznekoff and Rose, 2013). They also tend to be underrepresented in games as characters (Behm-Morawitz and Mastro, 2009), often depicted as weak, dependent ("damsels in distress"), and sexually objectified (Dill and Thill, 2007; Near, 2013), all of which creates an unwelcoming climate for women players. Despite the fact that women represent 45% of U.S. video game players (Entertainment Software Association, 2018), women are less likely to identify themselves as "gamers" and report less video game use than men (Crawford and Gosling, 2005). This further fuels the stereotype that women are not "real gamers" who prefer more casual and mobile games focused on fantasy and completion rather than action and competition (Hartmann and Klimmt, 2006; Yee, 2017).

Such gender stereotypes have a harmful, self-reinforcing effect (Kaye and Pennington, 2016; Shen et al., 2016) that potentially influences performance and attitudes not only in video game contexts but also in STEM contexts. Just as gender stereotypes introduced to children in the home (e.g., by parents) have been found to affect girls' beliefs about self-efficacy in STEM fields (Gunderson et al., 2012), gender stereotypes propagated through video games may have a similar effect, especially given that video games often serve as a gateway to STEM learning (Giammarco et al., 2015).

Combining these threads, this research examines the potential for gender stereotypes introduced in an AR or VR STEM-gaming context to influence stereotypic beliefs about women in STEM fields. This study is one of the first (Fordham et al., 2020) to examine the relationship between stereotypes in a STEM-gaming context and gender-stereotypic beliefs about STEM fields. This study also contributes a novel examination of differences in the outcomes of gender stereotypes between AR and VR in educational gaming. Such differences potentially relate to varying stereotypical associations with these two technologies. This study has implications for theoretical understandings of stereotypes in STEM-gaming contexts as well as for practical implications related to the development of AR and VR video games, particularly those that intend to promote equitable learning outcomes.

[1]Although that gender refers to non-binary identity characteristics in some contexts (Bem, 1981; Ansara and Berger, 2016; Lips, 2017), this paper focuses on gender-majority groups (i.e., men and women), consistent with video game literature (Kuznekoff and Rose, 2013; Kaye and Pennington, 2016; Shen et al., 2016) and previous research on gender-stereotypes (e.g., design, programming; Hargittai and Shafer, 2006; Nguyen and Ryan, 2008; Giammarco et al., 2015).

Stereotype Threat and Stereotype Reactance

In order to understand the effects of stereotypes in gaming contexts more deeply, we delve into concepts that explain how people respond to stereotypes, namely, stereotype threat and stereotype reactance. Stereotype threat occurs when individuals respond to a subtle reminder of a negative stereotype by conforming behaviorally to the stereotype (Steele, 1997). For example, a woman is more likely to perform poorly on a math test after being told that the test tends to yield gender differences—a subtle reminder of the stereotype—compared to being told that the test yields no gender differences (Spencer et al., 1999). This phenomenon occurs because subtle introductions of stereotypes have potent effects on a subconscious level (Nguyen and Ryan, 2008), leading to increased anxiety, arousal, and other factors that are often imperceivable to the individual (Shapiro and Neuberg, 2007). The phenomenon has been studied and replicated in many different contexts, from cognitive performance to physical activity to workplace performance (Azzarito and Harrison, 2008; Nguyen and Ryan, 2008; Pennington et al., 2016).

In contrast, stereotype reactance—derived from the theory of psychological reactance (Brehm, 1966)—occurs when individuals respond to a blatant reminder of a stereotype by behaving in ways that disconfirm the stereotype (Kray et al., 2004). For example, women were more effective in a negotiation task after being explicitly told that masculine traits are associated with negotiation success—an overt reminder of the stereotype—compared to receiving an implicit reminder of this stereotype. This phenomenon occurs when the recognition of a threat to individual freedom triggers anger and other negative emotions that lead the individual to attempt to assert their freedom (Miron and Brehm, 2006) and in a sense, resist stereotype threat (Pennington et al., 2016).

To summarize, when people experience stereotype threat, they conform to negative stereotypes about their social group. When people experience stereotype reactance, they act in ways that contradict the stereotype. The likelihood of whether someone experiences stereotype threat or reactance depends on whether the stereotype is triggered subtly or overtly, respectively. Thus, in the context of video games, when gender stereotypes are communicated subtly, they may have harmful effects (e.g., stereotype-consistent beliefs) through stereotype threat, but when made overt, these stereotypes may lead to stereotype reactance (e.g., counter-stereotypic beliefs).

Stereotypes, Video Games and STEM

The potential that stereotype threats in video game contexts influence performance and stereotypic beliefs—both within and beyond the gaming context—has been examined in multiple studies. In one experiment, a stereotype threat that prompted gaming ability led to women underperforming compared to men (Kaye and Pennington, 2016). In another study, when women participants who strongly identified as gamers were exposed to stereotype threat (i.e., a male-dominated leaderboard), they performed worse at a puzzle-platform game and reported lower self-confidence than women with weaker gamer identities (Vermeulen et al., 2016).

These studies support the idea that gender-related stereotype threats prompted in the context of educational video games influences behavior.

Given the potential connection between video game contexts and STEM fields, it is important to examine the relationship between gender stereotypes, gaming performance, and gendered-stereotypic beliefs about women in STEM fields. In one study (Fordham et al., 2020), women participants who read an article stating that men are more skilled than women at video games (compared to a women-are-as-good-as-men article) performed worse in a first-person-shooter game and rated STEM fields as better suited for men than women. Further, participants in that study who were led to believe that the opponent was a man compared to a woman also rated STEM fields as better suited for men than women. These findings are consistent with the notion that stereotype threats in the context of video gaming impact women, illustrating the link between stereotypes in video game contexts and in STEM fields.

In a second study (Fordham et al., 2020), participants were presented either a hateful nonsexist or sexist message, prompting experiences of threat and sexism. Afterward, they customized a video game character that would represent either the shooter game's story or themselves, prompting self-concept, which was expected to increase the recognition of the stereotype threat and thereby trigger reactance. The condition that made the stereotype most blatant—a sexist message plus a self-representing avatar—led to more positive beliefs about women in STEM fields, supporting the expectation of stereotype reactance. In contrast, the condition that made the stereotype subtle—a sexist message plus a game-representing avatar—led to more negative beliefs about women in STEM fields, consistent with stereotype threat. This study supports the notion that stereotype threat and stereotype reactance outcomes in video game contexts—reflected by stereotypic beliefs about women in STEM fields—depend on the extent to which the gender and gaming stereotype is made subtle or overt.

Together, these findings suggest that reminders of gender stereotypes in gaming contexts may influence game performance and the endorsement of gender stereotypes in STEM fields, but the blatancy of the reminder influences whether the stereotype has negative (stereotype threat) or positive (stereotype reactance) outcomes. In this largely understudied context of video games and STEM stereotypes, it is difficult to predict whether a stereotype reminder is blatant enough to exceed the threshold of stereotype reactance. If the prompt is below this threshold, then we would expect it to cause stereotype threat; namely, subsequent performance and beliefs would be aligned with the stereotype. If the prompt is overt enough to trigger stereotype reactance, then we would expect subsequent performance and beliefs to contradict the stereotype. In either case, we expect that people who have been prompted with gender stereotypes prior to playing a STEM game would exhibit different performance compared to people who receive a counter-stereotypic prompt. However, because we do not know the extent to which the stereotype prompt will be perceived as subtle or overt, we pose a non-directional research question.

Research Question 1: Do gender stereotypes prompted in a STEM-gaming context influence 1) STEM-game performance, and 2) stereotypic beliefs about women in STEM fields?

Stereotypes, Augmented/Virtual Reality, and STEM

Augmented reality and virtual reality are becoming more commonplace in education contexts (Wu et al., 2013; Merchant et al., 2014). AR typically incorporates digital images into a physical space with real world objects while VR isolates the user from the physical environment and immerses them in a new virtual world (Milgram et al., 1995). In light of new advances, researchers have examined AR and VR effects on education and learning (Wu et al., 2013; Merchant et al., 2014). For example, VR and AR have been shown to differentially impact student learning (Huang et al., 2019). While such research provides valuable insights into the potential for these technologies to be integrated into educational contexts, few have examined how differences between these technologies relate to social factors, such as gender and stereotypes. There is a broad field of research on gender gaps in internet and computer use (Hargittai and Shafer, 2006; van Deursen and van Dijk, 2014) suggesting that gendered differences in self-efficacy and other attitudes about technology are associated with digital skills (Correll, 2001; Beckwith et al., 2007; Huffman et al., 2013). However, little if any research has examined such gaps with respect to AR and VR in educational contexts. Still, previous research on these two modalities can be synthesized to argue that AR and VR likely differ in the extent to which they are associated with gender stereotypes.

VR is Potentially More Gender-Stereotyped than AR

VR is potentially oriented toward and thus associated with men more than women, thereby reinforcing gender stereotypes related to this technology. Studies suggest men have an advantage over women in VR contexts, particularly regarding susceptibility to cybersickness, spatial tasks, and cognitive performance (Larson et al., 1999; Terlecki and Newcombe, 2005; Munafo et al., 2017). The aptly named study Virtual Reality Is Sexist: But It Does Not Have to Be (Stanney et al., 2020) found that incorrect interpupillary distance (IPD) fit of head-mounted VR devices contributes to sex differences in cybersickness, suggesting that VR devices have not been designed to sufficiently consider women users. Although a systematic review found conflicting evidence of sex differences in cybersickness (Grassini and Laumann, 2020)—and other sex-associated technology-use differences such as spatial rotation skills-have been found to dissipate after practice (Rodán et al., 2016; Spence and Feng, 2010), cultural assumptions and stereotypes about this technology have persisted. Men are more likely to own and intend to own consumer VR headsets (Clement, 2021), possibly due to the growing availability and appeal of virtual reality video games (Foxman et al., 2020; Kosa et al., 2020), and men have been found to enjoy and intend to play virtual reality games more than women (Um, 2020).

In contrast, AR technologies are far less gender-stereotyped. The arguably most widely adopted AR game to date—Pokémon GO—has been popular across gender and age classifications (Serino et al., 2016), with one study receiving a higher response rate of women (58%) than men players from a sample of over 600 respondents recruited on internet forums. Further, in comparison to the offerings of VR games that are oriented toward men, AR games span a range of genres (Tan and Soh, 2010). AR is also being readily adopted outside of video gaming in more gender-balanced contexts, such as in social media and marketing (Bin Mohd Nasir, 2015). Together, AR seems to be evolving into a technology that is more balanced across genders and media applications than VR.

VR Masks Self-Identity Cues More than AR

In contrast with the argument that VR tends to be more male-stereotyped than AR, stereotype-triggering cues may become less salient during a VR than an AR task because VR is more likely to mask identity cues to a greater extent than AR (assuming there are no self-representing avatars in either scenario). In other words, VR isolates the individual from most reminders of the outside world, but when using AR, the individual can still see elements of the outside world. One such element of the outside world is the individual's own body, which is hidden from the user in VR, but is a direct representation of the individual's non-digital self in AR. Hence, in AR, users are more likely to be reminded of their own identity characteristics, including gender. In the context of gender stereotypes, VR users (compared to AR users) are less likely to be reminded of their own gender, making them less susceptible to gender-related stereotypes.

Do Stereotypes Effects Differ between VR and AR?

The previous two sections argue that VR might be stereotyped as male more than AR, but VR might also mask self-identity characteristics more than AR. These differences may influence the effects of a gender stereotype prompt within a STEM-gaming context (e.g., reading an article stating that women perform worse with digital technology than men compared to an article saying they perform equivalently). As described earlier, if a stereotype is presented in a subtle way, it is more likely to lead to stereotype threat than if it is presented in a blatant way, which potentially leads to stereotype reactance.

Differences between VR and AR in gender-stereotype associations and in the salience of self-identity characteristics may influence perceptions of a stereotype prompt as being subtle or blatant, thereby influencing whether there is a stereotype threat or stereotype reactance response. If VR is associated with gender stereotypes to a greater extent than AR is, a gender-stereotypic prompt in a VR context could make the gender stereotype more blatant than it would be in an AR context. If this prompt then exceeds the threshold to be perceived as blatant—which would be more likely in VR than AR—then it would lead to stereotype reactance in VR and stereotype threat in AR. At the same time, if VR masks identity cues to a greater extent than AR does, a gender-stereotypic prompt in a VR context might be perceived as more subtle than it would be in an AR context, where users are more likely already primed with a gender-associated self-concept. If this prompt then exceeds the threshold to be perceived as blatant—which would be more likely in AR than VR—then it would lead to stereotype reactance in AR and stereotype threat in VR. Hence, because we do not know the extent to which a stereotype prompt will be perceived as subtle or overt, we pose the following open-ended question.

Research Question 2: Is the effect of gender-stereotype prompts in a STEM-gaming context on 1) STEM-game performance and 2) beliefs about women in STEM fields is moderated by modality (i.e., AR or VR)?

Performance as Mediator of Effects on Stereotype-Consistent Beliefs

Up until this point, effects on STEM-gaming performance and gender-stereotypic beliefs about STEM fields have been discussed as separate outcomes of gaming-related stereotype reminders, but the two outcomes may be related. If performance is interpreted as an indicator of ability, then negative performance in the STEM-gaming context may lead individuals to believe that they are not as well-suited for related contexts, such as STEM fields. Thus, the harmful effect of gaming-related stereotype threat on beliefs about women in STEM fields—reflected by attitudes about how well women perform relative to men in those specific fields (e.g., science, technology, etc.) might be mediated by the harmful effect of gaming-related stereotype threat on STEM-gaming performance. Taking the previous expectation that modality moderates the effect of stereotype prompts, we hypothesize the following moderated mediation relationship.

Hypothesis 1: The effect of gender-stereotype prompts on beliefs about women in STEM fields—moderated by gaming modality—is mediated by STEM-game performance.

Effects on Non-STEM Fields?

One final consideration is whether the effects of stereotype threats in a STEM-gaming context is restricted to STEM fields or if they extend into non-STEM fields as well. The arguments here suggest that stereotypical associations of gaming contexts relate more to STEM fields than non-STEM fields given the stronger technical focus of both video games and STEM. However, it is possible that the effect of stereotype threat in a STEM-gaming context also extends into non-STEM fields through other associations, such as changes in mood, self-efficacy, or motivation. Therefore, we are interested to see if the effect of the gender-stereotype prompts extends into participants' beliefs concerning how well women perform relative to men in non-STEM fields (e.g., English, Language, Education, and Humanities). In order to put this logic up against the claim that the phenomenon is unique to STEM fields, we pose the following question.

Research Question 3: Do the effects of stereotype threat on beliefs about women in STEM fields also extend into non-STEM fields?

MATERIALS AND METHODS

Participants

The current study builds off the same dataset as a previously published paper (Huang et al., 2019), which focused on different aspects of the dataset (i.e., spatial presence and knowledge-retention differences between AR and VR, but not gender stereotypes). To explore the given hypotheses, the current study consists of a between-subjects design with participants randomly assigned to one of the conditions in the 2 (modality: AR or VR) x 2 (stereotype prompt: stereotype-reinforcing or counter-stereotype). Participants from a large Midwestern university took part in an Institutional Review Board approved study. Through an interdepartmental research subject pool, 109 participants, with an average age of 20.5 years old ($SD = 1.61$), were recruited for extra credit. Given the focus of the study on gendered stereotype threat and STEM-related beliefs, we chose to include only women participants that passed the manipulation check in the subsequent analyses ($n = 64$).

Materials

The current study was conducted in a 10′ x 8′ segmented office space in order to reduce distractions. The office contained beige curtains, beige walls, and an empty desk with only a desktop computer. The desktop computer contained the survey that was already projected on the screen when the participant entered.

The Solar System–Space Museum mobile app, developed by ZeeMelApps (available at https://play.google.com/store/apps/details?id=com.zeemel.spacemuseumvr), was displayed on a Samsung S4 smartphone. This app allowed for the digital and auditory content to be presented in both VR and AR modes while still providing the same amount of information. The app showed three-dimensional visual representations of the solar system complimented by auditory commentary and information about the objects shown in the app such as planets and galaxies. In the AR mode, participants viewed the physical environment around them with a non-interactive layer on top of the portrayed image of the solar system digital content. Participants in this AR mode held the smartphone in their hands. In the VR mode, participants wore a Mattel VR Viewmaster phone-based VR headset. The solar system digital content in the VR mode was displayed in front of a white background.

Although the navigation screens were different in the two modes, researchers explained how to select the same option in both modes to participants to reduce discrepancies. Specifically, in the AR mode participants touched the "solar system" option on the screen, whereas in the VR mode participants used gaze selection to highlight "solar system" and then selected the highlighted option by pressing on a button on the top of the VR headset. To increase consistency between the two modes, participants were told to stand up and were explicitly told they were able to move in 360° to view the digital content in both modes. Participants were additionally reminded in the AR mode to hold the smartphone in front of their faces. The audio content transmitted through the smartphone's built-in speakers was identical in each mode.

Procedure

After signing a consent form which provided a brief description of the research purpose (i.e., to understand how different people experience mixed reality games), participants were directed to a desk that projected a survey on the screen. Participants completed a pre-test questionnaire that contained questions measuring their solar system knowledge, demographics, and were randomly prompted to read one of two short articles that presented an abstract from what appeared to be a published research article. The fictitious article abstracts, derived by the research team by making minor changes to an actually published abstract (Shen et al., 2016), contained a stereotype-reinforcing or a counter-stereotype statement about women's performance and participation in digital technologies. The stereotype-reinforcing article (see **Supplementary Appendix A**) claimed men advanced faster than women in technology use, whereas the counter-stereotype article (see **Supplementary Appendix B**) said men and women advance equally. At the end of the survey, participants answered a question about the study presented in the article as a manipulation check.

At the conclusion of the pre-survey, participants were instructed on the screen to tell the research assistant that they had completed the survey. The research assistant then explained how to use the solar system application (either AR or VR, depending on condition) and to pay attention to the information presented on the app. The research assistant then situated the participant with the technology, asked for questions, and gave instructions on how to start the app. After the participant selected the "solar system" option on the app, visual and auditory information played on the app for under 5 min. This time allowed for the "solar system" option to completely display the entirety of the visual and audio information. Participants then played a short game in which they destroyed incoming asteroids by moving their heads to aim and shoot (within the Solar System app) or by using their finger on the phone screen for 2 min—in order to reinforce the gaming-related nature of the context—and then, completed the post-survey, starting with items on performance, then spatial presence, and concluded with items on gender-stereotypic beliefs about STEM and non-STEM fields.

Measures

Performance was measured using an original scale developed for this specific context. Past research created similar indices to gauge performance by using learning outcomes in both AR (Lin et al., 2013) and VR (Kockro et al., 2015). Participants were aware that their performance assessment would be based on information presented in the game. Therefore, they were asked multiple choice questions about the information provided by the mobile app either through a spoken voice (e.g., "When did the solar system form?") or visually on the screen (e.g., "What was the color of Neptune?"). Given that learning efficacy for auditory and visual information differs between AR and VR (Huang et al., 2019), we included an even mix of information provided through visual and auditory channels in our analysis. In order to ensure a sufficient variance in performance between participants, items were only retained if more than 30% of participants answered correctly. In other words, the items which 70% or more participants incorrectly were deemed too difficult—potentially leading to a floor effect and noise in the data—so they were not included in the composite metric, thereby ensuring that this metric had a sufficient level of variance to reflect a signal in the data. Our final measure of performance included five audio questions and five visual questions. Participants' performance score was computed as the proportion of questions answered correctly (M = 0.60, SD = 0.21).

Gender-stereotypic beliefs about STEM fields was derived from ratings for each of five fields—Science, Computer Science, Engineering, Mathematics, and Video Game Design—in response to the question, "please rate how much you think men or women are better at the following." Responses were coded on a 100-point (unnumbered) sliding scale anchored by "WOMEN are much better" and "MEN are much better", with higher scores indicating more positive beliefs about women. A composite score was generated from the mean response across the five fields, measured in both the posttest (α = 0.86; M = 41.89, SD = 13.62)—which was a primary dependent variable of interest—and pretest (α = 0.88; M = 41.75, SD = 15.90) as a covariate used to control for the potential influence or pre-existing gender-stereotypic beliefs about STEM fields.

Gender-stereotypic beliefs about non-STEM fields was derived from responses to the same question as in the STEM-fields measure, but with respect to these fields: Humanities, English, Education, and Language. A composite score was generated from the mean response across the five fields, measured in both the posttest (α = 0.89; M = 64.75, SD = 15.46) and pretest (α = 0.88; M = 65.36, SD = 15.81), as with the previous measure. The questionnaire interspersed the STEM and non-STEM items in order to help mask the study purpose.

Spatial presence—the perceptual illusion of physically being in a mediated space (Biocca, 1997; Lombard and Ditton, 1997)—differs between AR and VR and this may be an important cause of differences in the outcomes of using these media modalities (Riva et al., 2016). One study found that using the VR (fully mediated) compared to AR (digital overlay onto a physical environment) mode of the same educational application led to more spatial presence, which mediated the effect on application-related task performance (Huang et al., 2019). Hence, spatial presence was included as a covariate in order to control for its potential effect on learning performance and stereotypic attitudes in this context. This was measured with five items from a revised version of an immersive virtual technologies scale on a 7-point Likert scale (Fox et al., 2009). Example items include "To what extent did it feel like you visited another place?" and "To what extent did you feel surrounded by the virtual world?". A composite measure was constructed from a mean of these scores (α = 0.88).

Space Knowledge was measured on the pretest with ten multiple choice questions regarding general knowledge about our solar system (e.g., "Is Earth larger or smaller than most of the other planets?", "Which are the gas giants?", and "True or false: the Sun's gravity is the strongest gravity in the solar system.") Each participant was given a point for each right answer, then these points were added together for a final space knowledge score (M = 5.38, SD = 1.54). This measure was included as a covariate given the potential that space knowledge prior to the study would influence participants ability on the main performance measure.

RESULTS

Manipulation Check and Pretest Equivalence

A manipulation check was employed to ensure participants read and understood the stereotype-reinforcing or counter-stereotype article through answering the question, "According to the article, women advance ____ in skill level as men." Out of 81 participants, 17 answered incorrectly, likely because they did not carefully read or understand the abstract, which was in academic language. Given that it was essential for participants to understand the article, those who failed the manipulation check were removed from the analysis, leaving 64 participants.

To confirm that scores for pretest variables did not differ between condition, we conducted a multivariate analysis of variance (MANOVA) test with stereotype condition and AR/VR condition as the fixed factors and the three pretest outcomes of interest (gender-stereotypic beliefs about STEM fields; gender-stereotypic beliefs about non-STEM fields; space knowledge) as the dependent variables. Neither the main nor interaction effects were found to be significant for the omnibus multivariate test (all p values over 0.680) nor for the individual between-subjects tests (all p values over 0.234), suggesting that random assignment to condition was successful.

Analysis of Covariance Tests on STEM-Game Performance

In order to examine RQ1a (Do gender stereotypes prompted in a STEM-gaming context influence STEM-game performance?) and RQ2a (Is the effect of gender-stereotype prompts on STEM-game performance moderated by modality?), we conducted an analysis of covariance (ANCOVA) with stereotype condition and AR/VR condition as the manipulated independent variables, spatial presence and space knowledge as covariates, and participant performance as the outcome variable. No main effects of the manipulated independent variables were significant, but a significant interaction effect was found for stereotype threat by AR/VR condition, F (1, 63) = 12.05, p < 0.001, η_p^2 = 0.17 (see **Figure 1**). For participants in the AR condition, the stereotype-reinforcing article hindered STEM-game performance (M = 0.51, SE = 0.05) compared to counter-stereotype article (M = 0.65, SE = 0.05), but for participants in the VR condition, the stereotype-reinforcing article was associated with better performance (M = 0.72, SE = 0.05) compared to counter-stereotype article (M = 0.53, SE = 0.05).

To probe this interaction further, two simple effects tests were conducted. Analyzing only participants in the AR condition, those who received the stereotype-reinforcing article exhibited significantly worse performance (M = 0.54, $S.E.$ = 0.05) than those

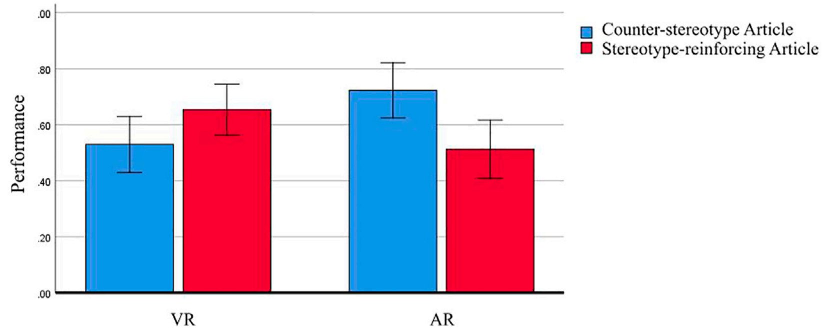

FIGURE 1 | Impact of stereotype threat prompt article and modality on performance (means and confidence intervals).

who received the counter-stereotype article (M = 0.75, $S.E.$ = 0.05), F (1, 29) = 10.00, p = 0.004, η_p^2 = 0.28, consistent with stereotype threat. Analyzing only participants in the VR condition, the difference approached significance, F (1, 33) = 3.55, p = 0.069, η_p^2 = 0.11, with those who received the stereotype-reinforcing article exhibiting better performance (M = 0.50, $S.E.$ = 0.05) than those who received the counter-stereotype article (M = 0.65, $S.E.$ = 0.05).

As a final probe of this interaction, two additional simple effects tests were conducted with splits in the opposite direction. Analyzing only participants in the stereotype-reinforcing condition, the difference approached significance, F (1, 31) = 4.02, p = 0.055, η_p^2 = 0.13, with those who used VR exhibiting better performance (M = 0.68, $S.E.$ = 0.05) than those who used AR (M = 0.52, $S.E.$ = 0.05). The difference among participants in the counter-stereotype condition was significant, F (1, 31) = 8.16, p = 0.008, η_p^2 = 0.23, with those who used VR exhibiting significantly worse performance (M = 0.50, $S.E.$ = 0.05) than those who used AR (M = 0.72, $S.E.$ = 0.05).

Therefore, these results inform RQ1a, suggesting that gender stereotypes do indeed influence STEM-game performance. Modality was found to moderate the effect of gender stereotype on performance. According to the simple-effects tests, the stereotype-reinforcing article hindered performance in AR (consistent with stereotype threat) but increased performance in VR (consistent with stereotype reactance, albeit approaching significance). Further, for participants who read the counter-stereotype article, VR was associated with significantly worse performance than AR, while for participants who read the stereotype-reinforcing article, VR was associated with better performance than AR (albeit approaching significance). Together, these results inform RQ2a, suggesting that the stereotype-reinforcing prompt led to stereotype threat in AR and stereotype reactance in VR.

Analysis of Covariance Tests on Gender-Stereotypic Beliefs About STEM

In order to examine RQ1b (Do gender stereotypes prompted in a STEM-gaming context influence gender-stereotypic beliefs about STEM?) and RQ2b (Is the effect of gender-stereotype prompts on beliefs about women in STEM fields moderated by modality?), we conducted a repeated measures ANCOVA with stereotype condition and AR/VR condition as the manipulated independent variables, spatial presence and space knowledge as covariates, and change in gender-stereotypic beliefs about STEM from pretest to post-test as the outcome variable. No main effects of the manipulated independent variables were found significant, but a nearly significant interaction effect was found for stereotype threat by AR/VR condition, F (1, 58) = 4.00, p = 0.051, η_p^2 = 0.06 (see **Figure 2**). For participants in the AR condition, the stereotype-reinforcing article was associated with a negative change in beliefs about women in STEM fields (M = −1.95, SE = 2.56), while the counter-stereotype article was associated with a positive change in beliefs (M = 1.28, SE = 2.41). Conversely, for participants in the VR condition, the stereotype-reinforcing article was associated with a positive change (counter-stereotypic) in beliefs about women in STEM fields (M = 3.65, SE = 2.25), while the counter-stereotype article was associated with a negative change in beliefs (M = −2.51, SE = 2.45).

To probe this interaction further, two simple effects tests were conducted. Analyzing only participants in the AR condition, no significant difference was found between the stereotype-reinforcing article and counter-stereotype article conditions. Analyzing only participants in the VR condition, there was a significant difference, F (1, 30) = 4.25, p = 0.048, η_p^2 = 0.12, with those who received the stereotype-reinforcing article exhibiting more positive change (counter-stereotypic) in beliefs about women in STEM fields (M = 4.15, SE = 1.97) than those who received the counter-stereotype article (M = −1.93, SE = 1.97).

As a final probe of this interaction, two additional simple effects tests were conducted with splits in the opposite direction. Analyzing only participants in the stereotype-reinforcing condition, there was a significant difference, F (1, 28) = 4.30, p = 0.048, η_p^2 = 0.13, with those used VR exhibiting more positive change (counter-stereotypic) in beliefs about women in STEM fields (M = 3.20, SE = 1.70) than those in the AR condition (M = −2.17, SE = 1.82). Analyzing only participants in the counter-stereotype condition, no significant difference was found between the stereotype-reinforcing article and counter-stereotype article conditions.

These results inform RQ1b, with gender stereotypes prompted in the STEM-gaming context found to influence gender-stereotypic beliefs about STEM. Further, modality was found to

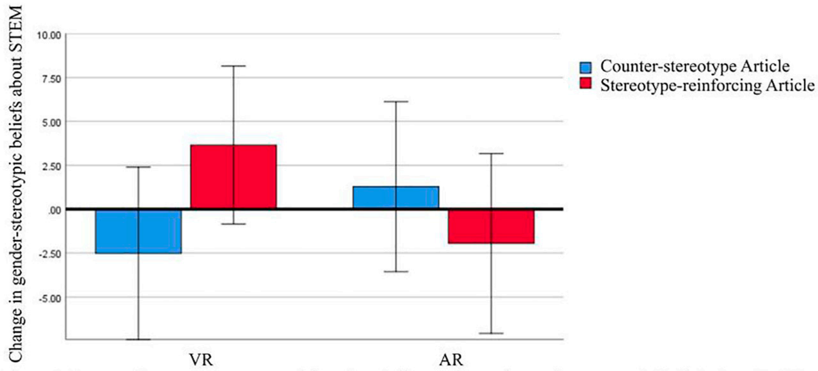

FIGURE 2 | Impact of stereotype threat prompt article and modality on change in gender-stereotypic beliefs about STEM fields (means and confidence intervals).

moderate the effect of gender stereotypes in the STEM-gaming context on gender-stereotypic beliefs about STEM. According to the simple-effects test, the stereotype-reinforcing article was associated with more counter-stereotypic beliefs in VR (consistent with stereotype reactance), though no difference was found in AR. Further, for participants who received the stereotype-reinforcing article, VR was associated with more counter-stereotypic beliefs than AR. Together, these results inform RQ2b and suggest that the gender-stereotype prompt induced stereotype reactance for participants in the VR condition.

Analysis of Covariance Test for RQ3

To examine RQ3 we conducted a repeated measures ANCOVA with stereotype condition and AR/VR condition as the manipulated independent variables, spatial presence and space knowledge as covariates, and change in gender-stereotypic beliefs about non-STEM fields from pretest to post-test as the outcome variable. No significant main effects nor interaction effects were found. These findings inform RQ3 (Do the effects of stereotype threat on beliefs about women in STEM fields also extend into non-STEM fields?), providing no evidence that gender stereotypes triggered in the STEM-gaming context influence perceptions of non-STEM fields.

Ordinary Least Squares Regression Path Analysis

Finally, we tested the expectation for moderated mediation in H1 (The effect of gender-stereotype prompts on beliefs about women in STEM fields—moderated by gaming modality—is mediated by STEM-game performance) by using ordinary least squares regression path analysis to perform a moderated mediation analysis (Hayes PROCESS, Model 7, 10,000 bootstrapped samples) with stereotypic beliefs about women in STEM (post-test) as the outcome, stereotype condition as the predictor, STEM-game modality as the moderator, STEM-game performance as the mediator, and stereotypic beliefs about women in STEM (at pretest), preknowledge, and spatial presence as covariates. The index of moderated mediation was significant [B = −4.12, LLCI: −8.60, ULCI −0.87], with STEM-game performance significantly mediating the effect of stereotype condition on stereotypic beliefs about women in STEM in the AR condition [B = −2.59, LLCI: −5.43, ULCI −0.53], but not in the VR condition [B = 1.53, LLCI: −0.19, ULCI 4.25]. In other words, for participants in the AR condition, the gender-stereotype prompt was associated with more negative attitudes about women in STEM fields and this effect was mediated by a reduction in STEM-game performance, but there was no such mediation effect for participants in the VR condition. These results provide partial support for H1.

DISCUSSION

In response to the potential link between gender disparity in video games and STEM fields, the present study examined whether gender stereotypes in a STEM-game influences gender-stereotypic views of STEM fields and whether such influence differs depending on gaming modality (i.e., AR vs. VR). Results suggest that gender stereotypes prompted before playing the STEM game—through an article reinforcing or countering gender stereotypes about gaming ability—influenced game performance and STEM beliefs in a direction consistent with stereotype threat in AR and with stereotype reactance in VR. Specifically, for participants who played in AR, the stereotype-reinforcing prompt (compared to the counter-stereotype prompt) was associated with worse game performance and more stereotype-consistent beliefs about women in STEM (albeit with the stereotype-consistent beliefs finding approaching significance in the simple-effects test). Conversely, for participants who played in VR, the stereotype-reinforcing prompt was associated with better STEM-game performance and more counter-stereotypic beliefs about women in STEM (albeit with the performance finding approaching significance). Further, the effect of the stereotype-reinforcing prompt on beliefs about women in STEM was mediated by game performance, though only in AR. In contrast, the stereotype-reinforcing prompt was not found to influence beliefs about non-STEM fields, likely because these fields are not stereotyped in the same way as video games and STEM fields. Altogether, these findings support the argument that stereotypes triggered in a STEM-gaming context have the potential to reinforce stereotypes in STEM fields. Further, VR games appear to be more likely than AR games to cause gendered stereotype threat in the absence of additional stereotype prompts, but AR games seem more likely to cause such threat than VR when additional stereotype reminders are present.

The differential effects found between AR and VR modality potentially results from the distinction between stereotype threat and stereotype reactance. For participants who played the game in AR, the stereotype-reinforcing (compared to counter-stereotype) prompt was associated with worse game performance (significant for both interaction and simple effects) and more negative beliefs about women in STEM fields (approaching significance for interaction effect). This aligns with stereotype threat: being reminded of the stereotype that women have lower gaming ability led these participants to conform to the stereotype through a variety of potential mechanisms (e.g., increased cognitive load, lower self-efficacy, etc.). In contrast, for participants who played the game in VR, the stereotype-reinforcing (compared to counter-stereotype) prompt was associated with better game performance (approaching significance in the simple-effects test) and more positive beliefs about women in STEM fields. This might have occurred due to stereotype reactance—the phenomenon that when a stereotype is made explicitly salient to a member of the stereotyped group, they resist and actively attempt to counteract it. In this case, VR might be more male-stereotyped than AR. Thus, the stereotype-reinforcing prompt with VR-based gameplay was like a double-dosage of the stereotype reminder to which the participants would have been more likely to experience stereotype reactance. In other words, the stereotype-reinforcing prompt was relatively subtle in the AR condition (leading to stereotype threat) and relatively explicit in the VR condition (leading to stereotype reactance).

The finding of mediation for AR users (i.e., stereotype-related article —> game performance —> stereotypic attitudes) supports the argument that game performance may serve as an attitude-reinforcing mechanism that fuels the vicious cycle of stereotypes in video game and STEM contexts. If performance is interpreted as an indicator of ability, then performing poorly in the gaming context may signal to individuals that they will also perform poorly in related contexts, such as STEM fields. Hence, for women, when stereotype threat hindered performance, it also reinforced stereotypic attitudes about women in STEM fields. Although this logic seems sound, future research should be used to confirm the pattern and delve deeper into the mechanisms, especially given that the pattern was only found for AR and not VR users. For example, although not supported by the present study, the argument leading up to RQ2 may relate to this question. Namely, in AR, users can still see elements of the outside world, including themselves, while VR isolates the individual from most reminders of the outside world. Hence, AR might communicate reminders of personal identity characteristics—such as gender—more than VR, potentially complementing other subtle stereotype reminders that trigger stereotype threat, such as game performance.

The finding that the stereotype-reinforcing prompt induced stereotype threat in AR (not VR) notwithstanding, the present results suggest that VR might be more likely to cause stereotype threat in general educational contexts where overt gender stereotype reminders are less common. This somewhat counterintuitive inference builds from the argument that VR is more male-stereotyped than AR—given gender differences in cybersickness, consumer adoption (Bin Mohd Nasir, 2015; Clement, 2021), and gaming (Serino et al., 2016; Um, 2020). Consistent with this logic, in the present study, the stereotype-reinforcing prompt in the male-stereotyped VR condition made gender stereotypes explicit, thereby triggering stereotype reactance. However, without such a stereotype-reinforcing prompt, the stereotypical association of the VR context alone seemed to be sufficient to trigger stereotype threat. Consistent with these two points, the simple effects tests found that for participants who received the stereotype-reinforcing prompt, participants in VR (compared to AR) exhibited higher STEM-game performance (approaching significance) and counterstereotypic STEM attitudes (significant). In contrast, for participants who received the counter-stereotype prompt, STEM-game performance was *lower* and counterstereotypic STEM attitudes were higher in VR than AR (considering the nearly significant interaction effect for the latter). One possible interpretation is that in the absence of a stereotype-consistent prompt, women who play a STEM game in VR are likely to perform worse and endorse more gender-stereotypic attitudes about STEM fields compared to those who play the same game in AR because VR is more male stereotyped (at baseline) than AR. An alternative (unexpected) interpretation is that the counter-stereotypical prompt causes AR users to experience stereotype boost, an improvement in performance after being exposed to a positive generalization about a personal social group (Shih et al., 2012), because identity cues such as gender are more salient in AR compared to VR. In other words, participants in AR were more likely than those in VR to be reminded of their gender identity—because VR literally occludes the user's view of their own body—and thus experience a (counter-stereotype) boost after reading an article saying that women are as strong as men at video games.

Implications

This study offers two fundamental implications: 1) VR is potentially more male-stereotyped than AR; 2) AR makes personal identity cues more accessible than VR. Regarding #1, an important practical implication is that educators implementing VR in learning contexts should consider that women and girls may perceive this technology to be less inclusive of them than men and boys do. Of course, these perceptions can only change through exposure, so instructors and educators should not shy away from using VR, but instead should actively work to mitigate stereotype threat by encouraging equal use of VR by women and girls and otherwise working to dispel any gender stereotypes about the technology. Regarding #2, practitioners should recognize that female users are likely more susceptible to stereotype threat when using AR compared to VR and thus should be careful to avoid any subtle stereotype cues about gender that might arise in the context.

Technology designers could consider these implications as well. Studies suggest that VR has a great potential to influence stereotypes and implicit biases through embodiment in avatars (Peck et al., 2013; Banakou et al., 2016; Christofi and Michael-Grigoriou, 2017; Farmer and Maister, 2017). Avatars can also likely be implemented into AR (e.g., seeing a digital filter on your body when you look in a real mirror) to similar effect, though the research on this is limited due to the increased technological complexity of developing functional products. In any case, avatars (and avatar design options/guidelines) can be designed to deemphasize stereotyped identity characteristics and this may help mitigate stereotype threat (Fordham et al., 2020). Designers and practitioners should consider such potential effects and implement avatars in VR and AR in deliberate ways that will minimize negative effects of stereotypes.

Limitations and Future Directions

Some important limitations of this study should be noted. Most notably, only women participants were included in the analysis due to the study's focus on gender-related stereotype threat. However, men's beliefs about women in STEM fields might also be influenced by reminders of gender stereotypes in gaming contexts. Future studies should include men participants given this potential effect of gender stereotyping in gaming contexts reinforcing gender stereotypes in STEM for men as well as women.

The sample size for this study was quite small, resulting in the study being somewhat underpowered. However, in studies of virtual reality, smaller sample sizes are not uncommon (Cummings and Bailenson, 2016). Further, according to a sensitivity analysis using G*Power (Faul et al., 2007) with this study's characteristics (i.e., sample size, degrees of freedom, number of covariates) as inputs, the study was able to detect outcomes with medium effect sizes ($f = 0.458$), and nearly all of the effects identified were in this range or larger. While future research should certainly use sample sizes that can provide sufficient power to detect smaller effects, this caveat should not detract from the reliability of the present findings.

The AR and VR technology utilized in this study was based on a mobile-phone platform in order to maintain consistency in the content between the experimental conditions; however, there are far more advanced AR and VR systems on the market that future research should explore. Further, the study relied on a single, education gaming context and this game was not particularly interactive. Hence, the results might have limited generalizability outside of educational gaming and with other game genres. Future research should compare AR and VR in other gaming (and also non-gaming) contexts. This is especially important because previous research has found some gender differences in gaming genre preferences (Greenberg et al., 2010), but these trends might be changing (Wohn et al., 2020) and stereotypes about gender differences in gaming ability are often inconsistent with reality (Shen et al., 2016; Ratan et al., 2020).

Finally, future research on this topic could add value with younger participants, implicit measures in addition to self-report, measures of social identities related to being a gamer (which might mediate effects of stereotype threat), assessments and controls of the participants' previous exposure to AR and VR technology, and qualitative methods (e.g., interviews) to better understand how women and men perceive the differences between AR and VR.

CONCLUSION

The present study adds to the growing evidence that just as video games have been touted as a means of improving gender equality in STEM fields, gender stereotypes in STEM-gaming contexts may contribute to gender inequity in STEM fields, which then reinforces the deleterious stereotypes across STEM-related contexts. Future research should continue to examine the factors that both fuel and could be harnessed to break this vicious cycle, such as game modality.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available due to the consent process used at the time of data collection. Requests to access the datasets should be directed to rar@msu.edu.

ETHICS STATEMENT

The study involved human participants and was reviewed and approved by the Institutional Review Board at Michigan State University. The participants provided their informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

RR: Conceptualization; Data curation; Formal analysis; Funding acquisition; Methodology; Project administration; Supervision; Roles/Writing – original draft; Writing – review and; editing JB: Conceptualization; Data collection; Formal analysis; Methodology; Project administration; Supervision; Roles/Writing – original draft; Writing – review and; editing SK: Formal analysis; Roles/Writing – original draft; Writing – review and; editing AG: Formal analysis; Writing – review and; editing K-TH: Conceptualization; Data collection; Methodology; Project administration; Supervision; Writing – review and; editing.

FUNDING

This research was partially supported by the Honors College at Michigan State University. The funding came in the form of a gift and the funders were not involved in the study design in any way. We would also like to thank the AT&T endowment to the Media and Information Department at MSU for supporting Dr. Ratan's AT&T Scholar position.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frvir.2021.636643/full#supplementary-material

REFERENCES

Ansara, Y. G., and Berger, I. (2016). Cisgenderism. in *The Wiley Blackwell Encyclopedia of Gender and Sexuality Studies*. Editors N. Naples, R. C. Hoogland, M. Wickramasinghe, and W. C. A. Wong (Oxford: Blackwell), 137–160. doi:10.1002/9781118663219.wbegss426

Azzarito, L., and Harrison, L., Jr. (2008). `White Men Can't Jump'. *Int. Rev. Sociol. Sport* 43 (4), 347–364. doi:10.1177/1012690208099871

Banakou, D., Hanumanthu, P. D., and Slater, M. (2016). Virtual Embodiment of White People in a Black Virtual Body Leads to a Sustained Reduction in Their Implicit Racial Bias. *Front. Hum. Neurosci.* 10, 601. doi:10.3389/fnhum.2016.00601

Beckwith, L., Inman, D., Rector, K., and Burnett, M. (2007). On to the Real World: Gender and Self-Efficacy in Excel. *IEEE Symposium on Visual Languages and Human-Centric Computing*. New York City, NY: VL/HCC 2007, 119–126.

Behm-Morawitz, E., and Mastro, D. (2009). The Effects of the Sexualization of Female Video Game Characters on Gender Stereotyping and Female Self-Concept. *Sex Roles* 61 (11-12), 808–823. doi:10.1007/s11199-009-9683-8

Bem, S. L. (1981). Gender Schema Theory: A Cognitive Account of Sex Typing. *Psychol. Rev.* 88 (4), 354–364. doi:10.1037/0033-295x.88.4.354

Bin Mohd Nasir, M. F. (2015). USING MOBILE AUGMENTED REALITY APPLICATION AS A MARKETING STRATEGY IN SOCIAL MEDIA. Available at: http://utpedia.utp.edu.my/15907/.

Biocca, F. (1997). The Cyborg's Dilemma: Progressive Embodiment in Virtual Environments [1]. *J. Computer-Mediated Commun. JCMC* 3 (2). doi:10.1111/j.1083-6101.1997.tb00070.x

Brehm, J. W. (1966). *A Theory of Psychological Reactance*. Cambridge, MA: Academic Press.

Christofi, M., and Michael-Grigoriou, D. (2017). Virtual Reality for Inducing Empathy and Reducing Prejudice towards Stigmatized Groups: A Survey. 2017 23rd International Conference on Virtual System Multimedia (VSMM), Dublin, Ireland, October 31, 2017–November 2, 2017, 1– 8.

Clement, J. (2021). Virtual Reality (VR) and Augmented Reality (AR) Device Ownership and purchase Intent Among Consumers in the United States as of 1st Quarter 2017, by Gender. Statista. Available at: https://www.statista.com/statistics/740747/vr-ar-ownership-usa-gender/.

Collette, M. (2013). Games the Key to Girls' STEM Education. Available at: https://news.northeastern.edu/2013/06/28/girls-games/.

Correll, S. J. (2001). Gender and the Career Choice Process: The Role of Biased Self-Assessments. *Am. J. Sociol.* 106 (6), 1691–1730. doi:10.1086/321299

Crawford, G., and Gosling, V. (2005). Toys for Boys? Women's Marginalization and Participation as Digital Gamers. *Sociological Res. Online* 10 (1), 1–11. doi:10.5153/sro.1024

Cummings, J. J., and Bailenson, J. N. (2016). How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence. *Media Psychol.* 19 (2), 272–309. doi:10.1080/15213269.2015.1015740

Dill, K. E., and Thill, K. P. (2007). Video Game Characters and the Socialization of Gender Roles: Young People's Perceptions Mirror Sexist media Depictions. *Sex Roles* 57 (11-12), 851–864. doi:10.1007/s11199-007-9278-1

Entertainment Software Association (2018). Essential Facts about the Computer and Video Game Industry. Available at: http://www.theesa.com/about-esa/essential-facts-computer-video-game-industry/.

Farmer, H., and Maister, L. (2017). Putting Ourselves in Another's Skin: Using the Plasticity of Self-Perception to Enhance Empathy and Decrease Prejudice. *Soc. Just Res.* 30 (4), 323–354. doi:10.1007/s11211-017-0294-1

Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). G*Power 3: a Flexible Statistical Power Analysis Program for the Social, Behavioral, and Biomedical Sciences. *Behav. Res. Methods* 39 (2), 175–191. doi:10.3758/bf03193146

Fordham, J., Ratan, R., Huang, K.-T., Zhou, W., and Silva, K. (2020). *Is Gender Disparity in STEM Fields Related to Gender Stereotypes in Videogames? an Experimental Examination of Stereotype Threat Context Transfer*. Newbury Park, CA: The American Behavioral Scientist.

Fox, J., Bailenson, J., and Binney, J. (2009). Virtual Experiences, Physical Behaviors: The Effect of Presence on Imitation of an Eating Avatar. *Presence: Teleoperators and Virtual Environments* 18 (4), 294–303. doi:10.1162/pres.18.4.294

Foxman, M., Leith, A. P., Beyea, D., Klebig, B., Chen, V. H. H., and Ratan, R. (2020). Virtual Reality Genres: Comparing Preferences in Immersive Experiences and Games. *Extended Abstracts of the 2020 Annual Symposium on Computer-Human Interaction in Play*. New York, NY: Association for Computing Machinery, 237–241.

Giammarco, E. A., Schneider, T. J., Carswell, J. J., and Knipe, W. S. (2015). Video Game Preferences and Their Relation to Career Interests. *Personal. Individual Differences* 73, 98–104. doi:10.1016/j.paid.2014.09.036

Granic, I., Lobel, A., and Engels, R. C. M. E. (2014). The Benefits of Playing Video Games. *Am. Psychol.* 69 (1), 66–78. doi:10.1037/a0034857

Grassini, S., and Laumann, K. (2020). Are Modern Head-Mounted Displays Sexist? A Systematic Review on Gender Differences in HMD-Mediated Virtual Reality. *Front. Psychol.* 11, 1604. doi:10.3389/fpsyg.2020.01604

Greenberg, B. S., Sherry, J., Lachlan, K., Lucas, K., and Holmstrom, A. (2010). Orientations to Video Games Among Gender and Age Groups. *Simulation & Gaming* 41 (2), 238–259. doi:10.1177/1046878108319930

Gunderson, E. A., Ramirez, G., Levine, S. C., and Beilock, S. L. (2012). The Role of Parents and Teachers in the Development of Gender-Related Math Attitudes. *Sex Roles* 66 (3), 153–166. doi:10.1007/s11199-011-9996-2

Hargittai, E., and Shafer, S. (2006). Differences in Actual and Perceived Online Skills: The Role of Gender*. *Social Sci. Q* 87 (2), 432–448. doi:10.1111/j.1540-6237.2006.00389.x

Hartmann, T., and Klimmt, C. (2006). Gender and Computer Games: Exploring Females' Dislikes. *J. Comp. Mediated Comm.* 11 (4), 910–931. doi:10.1111/j.1083-6101.2006.00301.x

Huang, K.-T., Ball, C., Francis, J., Ratan, R., Boumis, J., and Fordham, J. (2019). Augmented versus Virtual Reality in Education: An Exploratory Study Examining Science Knowledge Retention when Using Augmented Reality/Virtual Reality Mobile Applications. *Cyberpsychology, Behav. Soc. Networking* 22 (2), 105–110. doi:10.1089/cyber.2018.0150

Huffman, A. H., Whetten, J., and Huffman, W. H. (2013). Using Technology in Higher Education: The Influence of Gender Roles on Technology Self-Efficacy. *Comput. Hum. Behav.* 29 (4), 1779–1786. doi:10.1016/j.chb.2013.02.012

Jenson, J., de Castell, S., and Fisher, S. (2007). "Girls Playing Games: Rethinking Stereotypes," in Of the 2007 Conference on Future Play, Toronto, Canada, November 14–17, 2017. doi:10.1145/1328202.1328205https://dl.acm.org/doi/abs/10.1145/1328202.1328205

Kaye, L. K., and Pennington, C. R. (2016). "Girls Can't Play": The Effects of Stereotype Threat on Females' Gaming Performance. *Comput. Hum. Behav.* 59, 202–209. doi:10.1016/j.chb.2016.02.020

Kockro, R. A., Amaxopoulou, C., Killeen, T., Wagner, W., Reisch, R., Schwandt, E., et al. (2015). Stereoscopic Neuroanatomy Lectures Using a Three-Dimensional Virtual Reality Environment. *Ann. Anat. - Anatomischer Anzeiger* 201, 91–98. doi:10.1016/j.aanat.2015.05.006

Kosa, M., Uysal, A., and Eren, P. E. (2020). Acceptance of Virtual Reality Games. *Int. J. Gaming Computer-Mediated Simulations (Ijgcms)* 12 (1), 43–70. doi:10.4018/ijgcms.2020010103

Kray, L. J., Reb, J., Galinsky, A. D., and Thompson, L. (2004). Stereotype Reactance at the Bargaining Table: The Effect of Stereotype Activation and Power on Claiming and Creating Value. *Pers Soc. Psychol. Bull.* 30 (4), 399–411. doi:10.1177/0146167203261884

Kuznekoff, J. H., and Rose, L. M. (2013). Communication in Multiplayer Gaming: Examining Player Responses to Gender Cues. *New Media Soc.* 15 (4), 541–556. doi:10.1177/1461444812458271

Larson, P., Rizzo, A. A., Buckwalter, J. G., Van Rooyen, A., Kratz, K., Neumann, U., et al. (1999). Gender Issues in the Use of Virtual Environments. *CyberPsychology Behav.* 2 (2), 113–123. doi:10.1089/cpb.1999.2.113

Lin, T.-J., Duh, H. B.-L., Li, N., Wang, H.-Y., and Tsai, C.-C. (2013). An Investigation of Learners' Collaborative Knowledge Construction Performances and Behavior Patterns in an Augmented Reality Simulation System. *Comput. Educ.* 68, 314–321. doi:10.1016/j.compedu.2013.05.011

Lips, H. M. (2017). Sandra Bem: Naming the Impact of Gendered Categories and Identities. *Sex Roles* 76 (9), 627–632. doi:10.1007/s11199-016-0664-4

Lombard, M., and Ditton, T. (1997). At the Heart of it All: The Concept of Presence. *J. Computer-Mediated Commun. JCMC* 3 (2), 1–67. doi:10.1111/j.1083-6101.1997.tb00072.x

Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., and Davis, T. J. (2014). Effectiveness of Virtual Reality-Based Instruction on Students' Learning

Outcomes in K-12 and Higher Education: A Meta-Analysis. *Comput. Educ.* 70, 29–40. doi:10.1016/j.compedu.2013.07.033

Milgram, P., Takemura, H., Utsumi, A., and Kishino, F. (1995). Augmented Reality: a Class of Displays on the Reality-Virtuality Continuum. *Telemanipulator Telepresence Tech.* 2351, 282–292.

Miron, A. M., and Brehm, J. W. (2006). Reactance Theory - 40 Years Later. *Z. für Sozialpsychologie* 37 (1), 9–18. doi:10.1024/0044-3514.37.1.9

Munafo, J., Diedrick, M., and Stoffregen, T. A. (2017). The Virtual Reality Head-Mounted Display Oculus Rift Induces Motion Sickness and Is Sexist in its Effects. *Exp. Brain Res.* 235 (3), 889–901. doi:10.1007/s00221-016-4846-7

Near, C. E. (2013). Selling Gender: Associations of Box Art Representation of Female Characters with Sales for Teen- and Mature-Rated Video Games. *Sex Roles* 68 (3-4), 252–269. doi:10.1007/s11199-012-0231-6

Nguyen, H.-H. D., and Ryan, A. M. (2008). Does Stereotype Threat Affect Test Performance of Minorities and Women? A Meta-Analysis of Experimental Evidence. *J. Appl. Psychol.* 93 (6), 1314–1334. doi:10.1037/a0012702

Noonan, R. (2017). Women in STEM: 2017 Update. U.S Department of Commerce. Available at: https://eric.ed.gov/?id=ED590906.

Peck, T. C., Seinfeld, S., Aglioti, S. M., and Slater, M. (2013). Putting Yourself in the Skin of a Black Avatar Reduces Implicit Racial Bias. *Conscious. Cogn.* 22 (3), 779–787. doi:10.1016/j.concog.2013.04.016

Pennington, C. R., Heim, D., Levy, A. R., and Larkin, D. T. (2016). Twenty Years of Stereotype Threat Research: A Review of Psychological Mediators. *PLoS One* 11 (1), e0146487. doi:10.1371/journal.pone.0146487

Ratan, R., Shen, C., and Williams, D. (2020). Men Do Not Rule the World of Tanks: Negating the Gender-Performance Gap in a Spatial-Action Game by Controlling for Time Played. *Am. Behav. Scientist* 64 (7), 1031–1043. doi:10.1177/0002764220919147

Riva, G., Baños, R. M., Botella, C., Mantovani, F., and Gaggioli, A. (2016). Transforming Experience: The Potential of Augmented Reality and Virtual Reality for Enhancing Personal and Clinical Change. *Front. Psychiatry* 7. doi:10.3389/fpsyt.2016.00164

Rodán, A., Contreras, M. J., Elosúa, M. R., and Gimeno, P. (2016). Experimental but Not Sex Differences of a Mental Rotation Training Program on Adolescents. *Front. Psychol.* 7. doi:10.3389/fpsyg.2016.01050

Serino, M., Cordrey, K., McLaughlin, L., and Milanaik, R. L. (2016). Pokémon Go and Augmented Virtual Reality Games: a Cautionary Commentary for Parents and Pediatricians. *Curr. Opin. Pediatr.* 28 (5), 673–677. doi:10.1097/mop.0000000000000409

Shapiro, J. R., and Neuberg, S. L. (2007). From Stereotype Threat to Stereotype Threats: Implications of a Multi-Threat Framework for Causes, Moderators, Mediators, Consequences, and Interventions. *Pers Soc. Psychol. Rev.* 11 (2), 107–130. doi:10.1177/1088868306294790

Shen, C., Ratan, R., Cai, Y. D., and Leavitt, A. (2016). Do men advance Faster Than Women? Debunking the Gender Performance gap in Two Massively Multiplayer Online Games. *J. Comput-mediat Comm.* 21 (4), 312–329. doi:10.1111/jcc4.12159

Shih, M. J., Pittinsky, T. L., and Ho, G. C. (2012). Stereotype Boost: Positive Outcomes from the Activation of Positive Stereotypes. In *Stereotype Threat: Theory, Process, and Application* . (Ed.) M. Inzlicht, Vol. 320, pp. 141–156). Oxford University Press, xv.

Smith, J. L., Lewis, K. L., Hawthorne, L., and Hodges, S. D. (2013). When Trying Hard Isn't Natural. *Pers Soc. Psychol. Bull.* 39 (2), 131–143. doi:10.1177/0146167212468332

Spence, I., and Feng, J. (2010). Video Games and Spatial Cognition. *Rev. Gen. Psychol.* 14 (2), 92–104. doi:10.1037/a0019491

Spencer, S. J., Steele, C. M., and Quinn, D. M. (1999). Stereotype Threat and Women's Math Performance. *J. Exp. Soc. Psychol.* 35 (1), 4–28. doi:10.1006/jesp.1998.1373

Stanney, K., Fidopiastis, C., and Foster, L. (2020). Virtual Reality Is Sexist: But it Does Not Have to Be. *Front. Robotics AI* 7, 4. doi:10.3389/frobt.2020.00004

Steele, C. M. (1997). A Threat in the Air: How Stereotypes Shape Intellectual Identity and Performance. *Am. Psychol.* 52 (6), 613–629. doi:10.1037/0003-066x.52.6.613

Tan, C. T., and Soh, D. (2010). *Augmented Reality Games: A Review*. Eurosis: Proceedings of Gameon-Arabia. doi:10.1115/icone18-29100

Terlecki, M. S., and Newcombe, N. S. (2005). How Important Is the Digital divide? the Relation of Computer and Videogame Usage to Gender Differences in Mental Rotation Ability. *Sex Roles* 53 (5), 433–441. doi:10.1007/s11199-005-6765-0

Um, N. (2020). Do Gender, Types of VR Game, & Virtual Presence Matter in VR Games? *Int. J. Contents* 16 (4), 78–83. doi:10.5392/IJoC.2020.16.4.078

van Deursen, A. J., and van Dijk, J. A. (2014). The Digital divide Shifts to Differences in Usage. *New Media Soc.* 16 (3), 507–526. doi:10.1177/1461444813487959

Vermeulen, L., Castellar, E. N., Janssen, D., Calvi, L., and Van Looy, J. (2016). Playing under Threat. Examining Stereotype Threat in Female Game Players. *Comput. Hum. Behav.* 57, 377–387. doi:10.1016/j.chb.2015.12.042

Wiest, L. R., Sanchez, J. E., and Crawford-Ferre, H. G. (2017). *Out-of-school-time STEM Programs for Females: Implications for Research and Practice Volume I: Longer-Term Programs*. Charlotte, NC: IAP.

Wohn, D. Y., Ratan, R., and Cherchiglia, L. (2020). Gender and Genre Differences in Multiplayer Gaming Motivations. *Proc. HCI Int.* 12211, 233–248. doi:10.1145/3391614.3393653

Wu, H.-K., Lee, S. W.-Y., Chang, H.-Y., and Liang, J.-C. (2013). Current Status, Opportunities and Challenges of Augmented Reality in Education. *Comput. Educ.* 62, 41–49. doi:10.1016/j.compedu.2012.10.024

Yee, N. (2017). Beyond 50/50: Breaking Down the Percentage of Female Gamers by Genre. Quantic Foundry. Available at: https://quanticfoundry.com/2017/01/19/female-gamers-by-genre/.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Virtual Reality

ORIGINAL RESEARCH
published: 06 September 2021
doi: 10.3389/frvir.2021.678875

Check for updates

The Availability of a Hidden Real Reference Affects the Plausibility of Position-Dynamic Auditory AR

Annika Neidhardt and Anna Maria Zerlik*

Institute of Media Technology, Technische Universität Ilmenau, Ilmenau, Germany

This study examines the plausibility of Auditory Augmented Reality (AAR) realized with position-dynamic binaural synthesis over headphones. An established method to evaluate the plausibility of AAR asks participants to decide whether they are listening to the virtual or real version of the sound object. To date, this method has only been used to evaluate AAR systems for seated listeners. The AAR realization examined in this study instead allows listeners to turn to arbitrary directions and walk towards, past, and away from a real loudspeaker that reproduced sound only virtually. The experiment was conducted in two parts. In the first part, the subjects were asked whether they are listening to the real or the virtual version, not knowing that it was always the virtual version. In the second part, the real versions of the scenes where the loudspeaker actually reproduced sound were added. Two different source positions, three different test stimuli, and two different sound levels were considered. Seventeen volunteers, including five experts, participated. In the first part, none of the participants noticed that the virtual reproduction was active throughout the different test scenes. The inexperienced listeners tended to accept the virtual reproduction as real, while experts distributed their answers approximately equally. In the second part, experts could identify the virtual version quite reliably. For inexperienced listeners, the individual results varied enormously. Since the presence of the headphones influences the perception of the real sound field, this shadowing effect had to be considered in the creation of the virtual sound source as well. This requirement still limits test methods considering the real version in its ecological validity. Although the results indicate that the availability of a hidden real reference leads to a more critical evaluation, it is crucial to be aware that the presence of the headphones slightly distorts the reference. This issue seems more vital to the plausibility estimates achieved with this evaluation method than the increased freedom in motion.

Keywords: auditory augmented reality, binaural synthesis, six degrees of freedom, perceptual evaluation, plausibility, authenticity, internal reference

Edited by:
Missie Smith, Independent Researcher, Detroit, MI, United States

Reviewed by:
Sungyoung Kim, Rochester Institute of Technology, United States
Daniela Romano, University College London, United Kingdom

****Correspondence:***
Annika Neidhardt
annika.neidhardt@tu-ilmenau.de

Specialty section:
This article was submitted to Virtual Reality and Human Behaviour, a section of the journal Frontiers in Virtual Reality

Received: *10 March 2021*
Accepted: *21 July 2021*
Published: *06 September 2021*

Citation:
Neidhardt A and Zerlik AM (2021) The Availability of a Hidden Real Reference Affects the Plausibility of Position-Dynamic Auditory AR. Front. Virtual Real. 2:678875. doi: 10.3389/frvir.2021.678875

1 INTRODUCTION

Augmented Reality (AR) aims at adding virtual elements to the real environment (Azuma, 1997; Sicaru et al., 2018). Auditory Augmented Reality (AAR) describes the enrichment of a listener's actual environment with virtual sound sources or other virtual acoustic elements like reflectors or obstacles causing acoustic shadows. A common approach to realize AAR is to use dynamic binaural synthesis over headphones or hearables (Jot and Lee, 2016; Russell et al., 2016; Garí et al., 2019;

Nagele et al., 2021). In such reproduction, the position and orientation of the listener's head are tracked, and the headphone signals are adjusted by convolving the dry mono source signal with the corresponding binaural room impulse responses (BRIR) without a noticeable delay (Lindau, 2009; Brandenburg et al., 2020). A BRIR filter characterizes the transfer path of the sound from the sound source through the room to both ears of the listener or as a substitute head (and torso) simulator with microphones in the ears. BRIRs vary with the position and orientation of the source and receiver in the room. For consideration of source or listener motion, BRIR filters have to be updated regularly and rapidly (Neidhardt et al., 2018; Wefers and Vorländer, 2018). With the goal to realize such an AAR reproduction with low-cost devices (e.g., Heller et al. (2016)), there is the desire to identify the potential for optimization without affecting the quality of the resulting spatial auditory illusions. This process demands appropriate methods to evaluate the achieved quality. One essential question is how the created virtual acoustic object perceptually compares to the corresponding real version if there is a real version. In this context, Authenticity and plausibility have become important constructs.

According to Blauert (1997), Authenticity describes the agreement of the perceived acoustical scene with an external reference. Thus, a virtual acoustic object created with binaural reproduction is considered authentic if it cannot be distinguished from the corresponding real version in a direct comparison.

Slater (2009, 2018) has proposed the plausibility illusion as one of the key components in the perception of multi-modal VR realizations. He linked this term to the overall credibility of the scenario compared to a user's expectations. While sticking to this basic understanding, Kuhn-Rahloff (2011) has adopted the construct to evaluate acoustic reproductions. According to this proposal, plausibility describes the agreement of the perceived acoustic scene with the listener's internal reference. This internal reference is basically the expectation that results from a person's individual listening experience.

Latoschick and Wienrich (2021) have argued that in AR, "the central idea is to augment a physical space with additional computer-generated entities and not to artificially simulate a virtual space" [*p.* 5]. Rather than assuming an illusion of plausibility, like Slater (2009) and Skarbez et al. (2017), they have defined plausibility as "a state or condition during an XR experience that subjectively results from the evaluation of any information processed by the sensory, perceptual, and cognitive layers" [*p.* 5]. In addition, Latoschick and Wienrich (2021) have proposed a novel model describing XR experiences and effects wherein coherence and plausibility constitute central essential components. This model is still based on the idea that perceptual cues, sensory cues, and higher-order (cognitive) cues have to be in line with the experience and expectation of the user to achieve coherence and plausibility.

According to all these definitions, a virtual acoustic object is considered plausible if it fulfills the listener's expectations. Slater (2009) and Skarbez et al. (2017) have stated that a virtual element can be plausible even if the user knows it is not real. However, if a virtual replicate of a real sound object is in satisfactory agreement with the individual expectations of the listener, this listener will not be able to tell for sure that the acoustic object is virtual and will accept it as real. At this point, the highest degree of plausibility is achieved. If the internal reference is of limited accuracy, the listener may also accept an inaccurate virtual replicate as real. In contrast, listeners with a wrong internal reference may not even accept the real version as real. One of the challenges in evaluating plausibility is the limited reliability and stability of a listener's internal reference.

Several studies assessed the authenticity of spatial auditory illusions created with static binaural synthesis without the option of interactive listener motion (Moore et al., 2010; Maseiro, 2012; Oberem et al., 2016). Brinkmann et al. (2017) have presented the first study investigating the authenticity of virtual sound sources in different real rooms created with dynamic binaural synthesis considering interactive head rotation. For the realization, a simulated equivalent of a real scene is created based on individual BRIR measurements. For these measurements, extra-aural headphones (Erbes et al., 2012) were placed over the ears of the listener to consider their influence on listening to the real scene. An experiment with an individual two-alternative forced choice (2AFC) test paradigm was conducted to test for small noticeable differences. With the given realization, an authentic, dynamic binaural reproduction for interactive head rotation was achieved for the speech signal but not for the noise signal.

An authentic implementation demands high technical precision and effort. In AAR, usually, a direct comparison to the real version is not possible. Thus, for many applications, the concept of plausibility is more interesting. Lindau and Weinzierl (2012) have proposed a method based on the Signal Detection Theory to evaluate the plausibility of a dynamic binaural synthesis system. Again, a real sound field and its binaural simulation are considered. In the experiment, randomly, either the real scene or the binaural auralization was provided to the subjects. They had to decide in a Yes/No paradigm which version they were listening to. The basic idea of using a Yes/No paradigm in a mixture of real and virtual sound sources was not new at that point. This approach was employed, e.g., by Hartmann and Wittenberg (1996) to evaluate externalization and convincingness, by Langendijk and Bronkhorst (2000) to investigate the *fidelity* of virtual sound sources, and in an earlier study by Lindau et al. (2007). However, Lindau and Weinzierl (2012) have taken this approach to a new level of depth and linked it to plausibility, as proposed by Kuhn-Rahloff (2011).

Including a real sound source as a test case in an experiment requires considering how the presence of the headphones affects the perception of the real sound source. This effect is also added to the virtual version to avoid this occlusion or shadowing effect causes audible cues only for the real scene. A new set of BRIRs has to be measured with the desired pair of headphones placed on the listener's or the dummy's head. In the investigation of a 6DOF-system, this causes considerable effort because each position of interest has to be measured separately. Moreover, a slightly distorted perception of the real sound source caused by the occlusion can lead to additional confusion. On the one hand, listeners could increasingly mistake the real sound source for the virtual version. On the other hand, this approach can only investigate the quality of a spatial auditory illusion of a slightly distorted reality. This is a common challenge in realizing AAR systems, which provide virtual content alongside the real acoustic environment. Is it a suitable approach to encourage the creation of virtual content containing the same effect?

The method suggested by Lindau and Weinzierl (2012) is valid and interesting for evaluating the reproduction system itself. However, reproduction systems need to be tested for plausibility, as well as fictional scenes or other contents for which there are no real counterparts. If the scene contains a cartoon hero or a little ghost flying around or if a product is designed virtually and realized later on, how can we evaluate the plausibility in such cases? These questions are also interesting for Virtual Reality, where the listener can be transferred to a fantasy room like in the studies by Enge et al. (2020); Remaggi et al. (2019).

In the field of VR, scientists have started to distinguish between internal and external plausibility. Hofer et al. (2020) have provided a nice summary of that discussion. In this understanding, internal plausibility "refers to the extent to which the environment is consistent within itself or with respect to the expectations raised by its genre" [*p.* 2]. An example of violated internal plausibility, as defined by Hofer et al., would be to have a vegetarian that eats meat in the scene because the new information—the character eats meat—contradicts the already presented information—the character is a vegetarian. External plausibility in this context "refers to how consistent the virtual environment is to user's real-world knowledge" [*p.* 2]. This definition addresses whether the presented scenario could occur in the real world, but it is not necessarily indistinguishable from reality. These interpretations and classifications of plausibility refer to the credibility and consistency of the content rather than the rendering quality, which we consider in our discussion of plausibility. Our study only considers scenes that can occur in the real world, that is external plausibility as described by Hofer et al. Still, it is essential to note that methods to evaluate plausibility based on a comparison with a real counterpart have the limitation of not being helpful for fictional contents.

In three previous studies (Neidhardt et al., 2018; Kamandi, 2019; Neidhardt and Knoop, 2017), we have evaluated the plausibility of an interactive approaching motion towards a virtual sound source without considering a real scene. The participants were asked to rate plausibility directly with the four answering options "clearly plausible," "rather plausible," "rather not plausible," and "clearly not plausible." In all these studies, the position-dynamic binaural synthesis was realized with the same reproduction setup to create the spatial auditory illusion. Each study included at least one test case with a BRIR dataset fully measured in the corresponding room. In all studies, this fully measured scene was rated as plausible by all participants. Alongside plausibility, Neidhardt et al. (2018) and Kamandi (2019) have asked for continuity, externalization, sound source stability, and the impression of walking towards a sound source. In both experiments, the plausibility ratings varied substantially according to the degree of simplification of the selected test scenes. The results for plausibility show quite a strong correlation with all of the four other attributes. In contrast, for example, continuity and externalization, or externalization and sound source stability, exhibit very low correlation. This suggests that asking directly for plausibility provides a suitable evaluation of the overall impression of the spatial auditory illusion. Our previous studies provide meaningful evaluations of the plausibility of dynamic binaural walk-through scenarios, although no real counterpart was included in the test. However, we want to know how our system performs in an experiment taking the real version into account. Generally, it is of interest how the results of an evaluation in the two different paradigms compare. Would they lead to the same conclusion?

So far, it has not been investigated whether including a real sound field in the test paradigm would influence the result. If that is the case, it may be valuable to distinguish different kinds of plausibility, e.g., indicating the agreement with the pure internal reference or the tuned internal reference resulting from listening to the real version of the scenario. **Table 1** summarizes a selection of previous studies on the authenticity and the two proposed categories of plausibility of auditory illusions created with binaural technology. In addition, we ordered the studies by the considered degree of interactivity. In a static reproduction, no interactive motion is possible. Several studies already took interactive head rotation into account. The option to interactively walk to another position relative to the virtual sound source is still a quite new challenge concerning the evaluation of plausibility.

A potential tuning of the internal reference may occur in an indirect comparison with the real counterpart. Especially for AAR, the actual environment and its components are likely to influence the internal reference. Since the scenario allows for a direct comparison, maybe the term mixed reference is more appropriate in this case. Wirler et al. (2020) have proposed the concept of transfer-plausibility as the "ability of a virtualized source to stand alongside multiple real sound sources" and studied the plausibility of virtual sound sources in real environments under varying scene complexity in terms of the number of concurrent loudspeaker signals. The setup realized dynamic binaural synthesis with 6DOF, but the participants were seated during the experiment. Their results suggested that an increased scene complexity decreases the number of correctly identified virtual sound sources even with a rendering of lower quality. The concept of co-immersion proposed by Stecker et al. (2018) addresses this topic similarly.

It is likely that the number of sources or the scene complexity, as well as the type and the relative positions of the available real sound sources, influences the internal reference. If, for example, a virtual loudspeaker is created next to a real loudspeaker, achieving a quality of the illusion that listeners cannot identify as virtual may be more challenging than if the sound of a person riding a bicycle is added to an acoustic environment with a distant street full of cars.

With this new study, we want to evaluate our position-dynamic AAR system with the approach proposed by Lindau and Weinzierl. To our knowledge, this is the first time this

TABLE 1 | Summary of previous studies investigating plausibility and authenticity of binaural synthesis. Plausibility is split up into the two proposed categories of measuring the agreement with the pure internal reference or a tuned internal reference as a result of the indirect comparison with the real counterpart of the scene. This overview is not exhaustive but provides examples for each of the cases.

Binaural synthesis	Plausibility I pure internal reference	Plausibility II "tuned" internal reference	Authenticity external reference
Static reproduction	(✓)	Hartmann and Wittenberg 'Convincingness' (1996) Oberem et al. (2016), part B	Moore et al. (2010) Maseiro (2012) Oberem et al. (2016), part A
Head rotation	(✓)	Lindau et al. (2007) Lindau and Weinzierl (2012) Pike et al. (2014)	Brinkmann et al. (2017)
Rotation and translation	Neidhardt and Knoop (2017) Neidhardt et al. (2018)		
	This study, part I	This study, part II	

approach is applied to a system that provides interactive walking. Furthermore, it is of interest to estimate the relevance of including the real version in evaluating plausibility. Therefore, we created an experiment to assess the plausibility of the auditory illusions created with our AAR system with and without real versions of the scenes among the test items. The following section presents the technical realization of the evaluated AAR system, the test scenario chosen for the experiment, and the test design.

2 MATERIALS AND METHODS

The test scenario was realized in a seminar room of the university in Ilmenau. The participants had to wear headphones. The two loudspeakers standing in the room could reproduce sound either in reality or virtually over headphones. To create the virtual reproduction, BRIR measurements were conducted. The procedure is documented in this section. The test method demands measuring the BRIRs with headphones placed on the dummy's ears to consider their influence on the perception of the real sound field. This influence depends on the type of headphones.

2.1 Choice of Headphones

Satongar et al. (2015) have shown that the passive influence of headphones can cause spectral distortions, affect the effective interaural time difference, and reduce localization accuracy. Brinkmann et al. (2017) have used the extra-aural headphones BK211 presented by Erbes et al. (2012) for their experiment on authenticity. These headphones may be the best choice for a mixed-reality scenario with respect to the lowest impact of the headphone geometry on the perception of the real scene. However, the extra-aural headphones are quite large and heavy. They tend to move slightly on the head during motion despite all effort to attach them stably to the listener. It may be assumed that wearing these headphones does not allow for a perfectly natural motion. Especially during walking, people may move more carefully to avoid changing the headphone position on the head. For this reason, we decided not to use the extra-aural headphones in this experiment.

Lindau and Weinzierl (2012) and Pike et al. (2014) have used STAX headphones. These cover the ears completely and influence the sound reaching the ears from outside noticeably, for example, by damping the high frequencies. These occlusion or shadowing effects also depend on the direction of the sound incident. In an attempt to find a good compromise, AKG K1000 headphones with an opening angle of 45° on both sides were chosen for this experiment. These headphones are increasingly used for the realization of AR in general. They are less bulky than the extra-aural BK211 and still keep some space between their speakers and the listener's ears. **Figure 1** shows the setup. In the aftermath of this study, we analyzed these effects for different headphones, including all the mentioned ones (Schneiderwind et al., 2021). Our discussion considers these results.

2.2 Measurement of Binaural Room Impulse Responses

The seminar room chosen for this study has a size of 9.9 m × 4.7 m×3.1 m (volume V = 144 m^3) and a reverberation time T_{60} = 0.99 s (broadband). A G.R.A.S. Kemar 45BA with AKG K1000 headphones placed on the ears was set up on an electronic turntable Outline ET 250-3D at nine positions in 25 cm intervals along a line with a length of 2 m. Two loudspeakers, Genelec 1030A were positioned in the room, one in front of the line with a distance of 1.25 m to the closest position and one 1.25 m right of the line as illustrated in **Figure 2**. BRIRs were captured for an azimuth resolution of 2° over the full 360°. Elevation changes were not considered.

We ensured that the headphones did not move on Kemar's head while going through the different positions and head orientations during the measurement. After the BRIR measurement, the headphone transfer function (HpTF) was measured with the same placement of the AKG K1000. The headphone compensation filter was created from the measured HpTF following the least-squares approaches described by Schärer and Lindau (2009). The captured BRIRs and the created headphone compensation filter are provided as an open-access dataset by Neidhardt (2019).

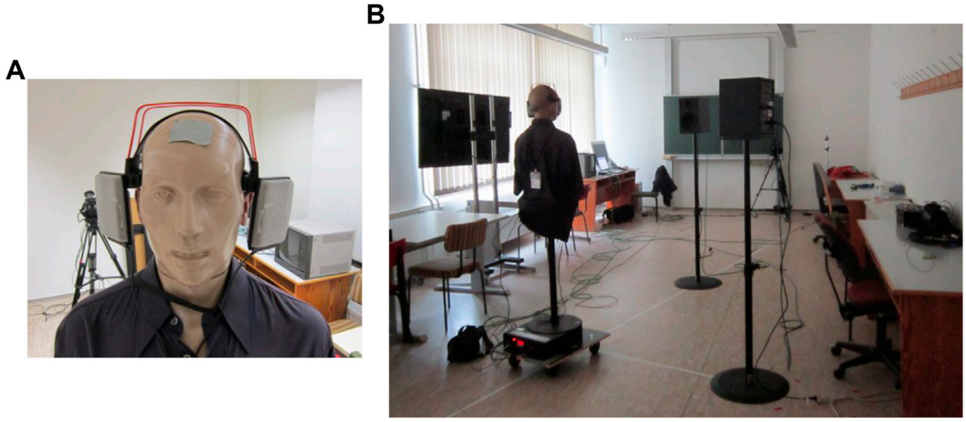

FIGURE 1 | (A) AKG K1000 headphones opened by 45° are placed on the Kemar 45BA's ears. **(B)** Setup for the BRIR measurement in the chosen seminar room.

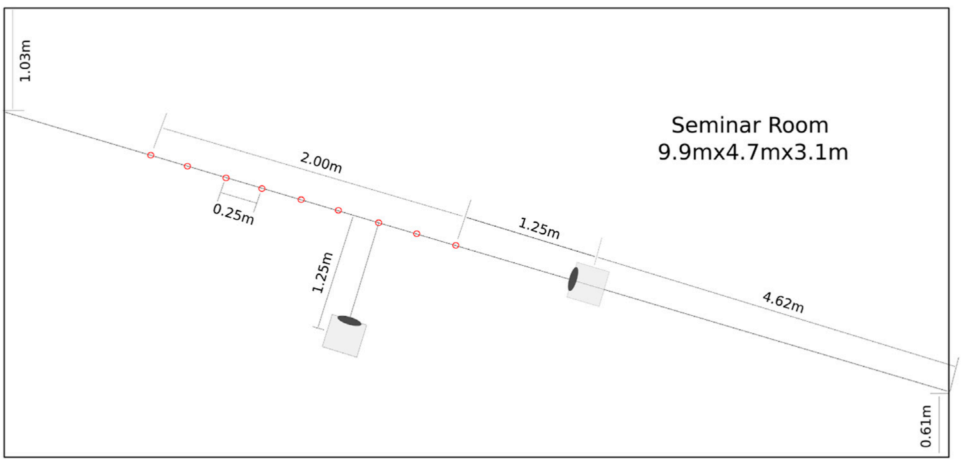

FIGURE 2 | Setup for the BRIR measurement in the chosen seminar room. AKG K1000 headphones were placed on the Kemar 45BA's ears throughout the measurement.

2.3 Position-Dynamic Reproduction Setup for Auditory AR

After the measurement, the two loudspeakers were kept in exactly the same positions of the same room. An HTC Vive tracker was attached to the headphones to track the position and orientation of the listener's head, as shown in **Figure 3**. The tracking module of the HTC Vive was calibrated to cover the area around the line of measured listening positions.

The Python tool *pyBinSim* presented by Neidhardt et al. (2017) was used for the partitioned convolution of the dry mono signal with the BRIR filters selected according to the tracking data. The filters had a length of 65,536 samples at a sampling frequency of 48 kHz. The block size was set to 512 samples. No interpolation or extrapolation was applied except for a cosine-square cross-fade in the time domain over the duration of one block size when switching to another filter. The real-time processing was executed by an Intel Core^{TM} i7-8700K (3.7 GHz) computer with 16 GB RAM and Windows 10 Enterprise (64-Bit). Audio reproduction was realized with an external sound card RME Fireface UCX. The sound level of the two reproduction setups was carefully adjusted by two expert listeners who compared both for several test stimuli.

2.4 Individualization of Binaural Audio

The BRIR filters used for dynamic binaural synthesis contain head-related information like interaural differences in level and time of arrival and spectral characteristics. These physical properties are important acoustic cues in spatial hearing and depend on the individual size and shape of a person's ears, head, and torso. They can vary substantially from person to person. If the binaural reproduction is based on head-related information that does not sufficiently match the listener's head, errors in sound source localization can occur and externalization can be affected. Both effects may reduce the overall quality of the auditory illusion in terms of plausibility. A wrong match of the individual ear distance can also cause instabilities of the perceived source position during motion. Thus, an individualization of the binaural reproduction is desirable but often requires considerable practical effort like individual BRIR measurements or at least a determination of individual interaural

FIGURE 3 | Test person wearing AKG K1000 headphones with a Vive tracker attached to them.

time difference (ITD) combined with an adequate BRIR adjustment. Brinkmann et al. (2017) have measured individual BRIRs for each participant before evaluating the authenticity of the binaural reproduction. Lindau and Weinzierl (2012) have conducted their study with two systems based on non-individual BRIRs measured with a FABIAN dummy head. In one of them, the ITDs were extracted and individually adjusted for each listener. With this system, a plausible reproduction according to the given test paradigm was achieved. For the other system, coloration and unstable localization were reported (Lindau et al., 2007). Pike et al. (2014) have tested the plausibility of dynamic binaural synthesis for head rotation with non-individual BRIRs of a small room with the method suggested by Lindau and Weinzierl (2012). The BRIRs were measured with a Neumann KU100 dummy, but an individualization of the ITDs was realized in the post-processing. Before the test, participants had to determine their ITD by listening to reproductions with different ITDs, which is not an easy task even for experts. With their setup, still slight instabilities in source localization were reported and described as increased localization blur or increased apparent source width. In their experiment, a sensory distance between real sound field and auralization was found.

In a test paradigm without considering a real scene, dynamic binaural synthesis with non-individual BRIRs was repeatedly perceived as plausible (Neidhardt and Knoop, 2017; Neidhardt et al., 2018; Kamandi, 2019).

2.5 Participants

Seventeen people aged between 18 and 33 years volunteered for participation in the experiment. The average age was 25 ± 2.57 years. Five of the subjects were experienced listeners in the field of BRIR-based binaural synthesis, and the others were mostly inexperienced. Experienced listeners were expected to be more critical about plausibility. For this reason, we were interested in recruiting at least a suitable number of them to allow for a separate analysis of this group. All participants were master students or Ph.D. candidates at the university in Ilmenau and interested in the field of AR. The selected group is considered representative of users of AR systems. The panel consisted of four female and 13 male listeners. All volunteers stated to have normal hearing abilities without any impairments. All participants completed the full experiment and all their results were included in the statistical analysis.

TABLE 2 | Two different source positions, three types of signals, and two different sound levels were taken into account.

Source position	Audio content	Gain
Frontal	Speech	0 dB
Frontal	Speech	–6 dB
Frontal	Music	0 dB
Frontal	Music	–6 dB
Frontal	Snare drum	0 dB
Frontal	Snare drum	–6 dB
Side	Speech	0 dB
Side	Speech	–6 dB
Side	Music	0 dB
Side	Music	–6 dB
Side	Snare drum	0 dB
Side	Snare drum	–6 dB

2.6 Test Scenes

The two different loudspeaker positions were considered as different test cases. Three test signals were included in the experiment:

- Speech: dry female speech reading an audiobook
- Music: pop song (left channel as mono)
- Snare drum: 50 bpm

Although the loudness of loudspeaker reproduction and binaural auralization was adjusted carefully, two different sound levels (0 dB and –6 dB) were included in the test to minimize the potential influence of minimal loudness differences in the determination process. This adds up to a total of 12 test scenes for each of the two reproduction methods. **Table 2** provides an overview.

All stimuli were band limited to a frequency range between 150 Hz and 16 kHz to reduce the influence of low-frequency background noise and loudspeaker distortion in the high frequencies.

2.7 Pre-Test with Few Experts

In the preparation of the official experiment, a few expert listeners conducted an informal pre-listening session. Both direct AB comparison and blind identification of auralization and real sound field were part of this procedure. The results and observations are documented in **section 3**. In the course of this critical listening session, the experts observed that a fade-in is required after activating the headphone reproduction. An abrupt start of the signal in the headphone reproduction revealed the virtual scene. This was considered in the final experiment.

2.8 Listening Experiment With the Test Panel

Before participating in the study, informed consent was obtained from all individual participants involved in the study. In the experiment, the participant had to wear the AKG K1000 headphones with the Vive tracker attached to them. At the beginning of each trial, the subject had to stand at the end of the translation line (measurement position with a distance of 3.25 m to the loudspeaker in the front). The participant was told that randomly either the real loudspeaker or its binaural simulation would be presented, and the task was to decide which of the two versions was currently active. In addition, the subject was instructed to move along the line and use head rotation and self-rotation arbitrarily.

The first part of the test aimed to investigate the plausibility with respect to the pure internal reference. For this part, it had to be avoided that the participant gets an impression of the real version of the sound field. Therefore, a training session was not feasible. In the second part, real scenes were included as test items to evaluate plausibility with regard to the internal reference tuned by the real versions of the scenes.

Test part I: All test scenes in their binaural version, 12 in total, were presented in a randomized order. The real reproduction was not included in this test. This part took about 15–20 min per participant.

Test part II: All test scenes in their binaural and their loudspeaker version, 24 in total, were presented in a randomized order. This part took about 30–40 min per participant.

The participants were asked to evaluate 36 test scenes wherein the number of virtual and real scenes is not necessarily similar. After the experiment, the participants were asked to describe the audible cues they used to distinguish simulation and real reproduction. The test procedure was designed in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

2.9 Required Sample Size and Test Duration

To achieve statistically meaningful results, an appropriate sample size is required. Furthermore, it is crucial to consider that taking time to explore the scene and take the decision may affect the rate of correct answers. Lindau et al. have conducted their experiment with 11 experienced listeners. Each of them had to evaluate 100 test samples. This allowed for an analysis of the individual sensitivity d_i', hence, the discriminability, based on the Signal Detection Theory. However, in their experiment, each test stimulus had a duration of only 6 seconds, which was possible because interactive self-motion was limited to ±80° in azimuth. The test was restricted to a one-time listening per sample. The authors reported that none of the participants took longer than 15 min for the whole test.

In our experiment, each of the 17 participants completed 36 evaluations. The participants were allowed to listen to and explore the scene as long as they thought it was helpful. On average, the assessment took the subjects 70 s per test scene. Between the scenes, there was a break of 20–25 s for the test conductor to take notes and start the new scene. In total, the experiment with introduction and interview at the end took between 50 and 70 min. Due to the breaks, the active exploration of the scene, and the reportedly interesting task, listener fatigue was kept at an acceptable level.

Especially in systems with a high degree of interactivity, there will always be a trade-off between a large sample size and providing the participants a suitable amount of time to explore the scene and make their decisions.

TABLE 3 | Possible outcomes in the Signal Detection paradigm.

Response	Virtual (signal)	Real (noise)
"Virtual"	Hit	False Alarm (FA)
"Real"	Miss	Correct Rejection (CR)

2.10 Methods for Statistical Analysis

A standard method to analyze the results of an experiment conducted in a Yes/No paradigm is based on the Signal Detection Theory. The following paragraph explains how the SDT can be used to estimate the discriminability between real and virtual reproduction.

2.10.1 Estimating the Discriminability Based on Signal Detection Theory

The participants have two answering options, "virtual" and "real." The type of reproduction can also be both virtual or real. If the participant cannot detect a cue indicating that the virtual sound source is active, the participant is more likely to pick the answer "real." Based on this idea, the real sound source is regarded as "Noise" and the virtual sound source with potential revealing cues as "Signal." In accordance with Herzog et al. (2019), the four possible outcomes in this classic SDT experiment are called Hit, Miss, False Alarm, and Correct Rejection. **Table 3** provides an overview.

The primary goal of SDT is to determine the sensitivity index d' and the decision criterion c. In this specific case, d' is the sensitivity to cues revealing the virtual reproduction as virtual. Thus, a sensitivity $d' = 0$ indicates that the virtual sound source cannot be distinguished from the real sound source. In this case, "perfect plausibility" would be achieved. The sensitivity is a measure for the discriminability of the virtual sound source from the real one. The decision criterion indicates whether there are any tendencies towards one of the two answers.

Using SDT, the most consistent analysis is possible if one observer completes many assignments for the same stimulus in its virtual and real version. If more subjects and more stimuli are taken into account, the theory demands determining the individual sensitivity d' for each combination of subject and stimulus separately and then calculating the mean sensitivity. If the sample size for each combination is too small, the sensitivity has to be determined for a pool of observers and stimuli. This pooled sensitivity is discussed in detail in Macmillan and Creelman (2004) [*p.* 331 ff].

Several previous plausibility studies have used the SDT for their analysis. For example, Lindau and Weinzierl (2012) have calculated the individual sensitivity per person, averaging over different signals and source positions, then calculated the mean

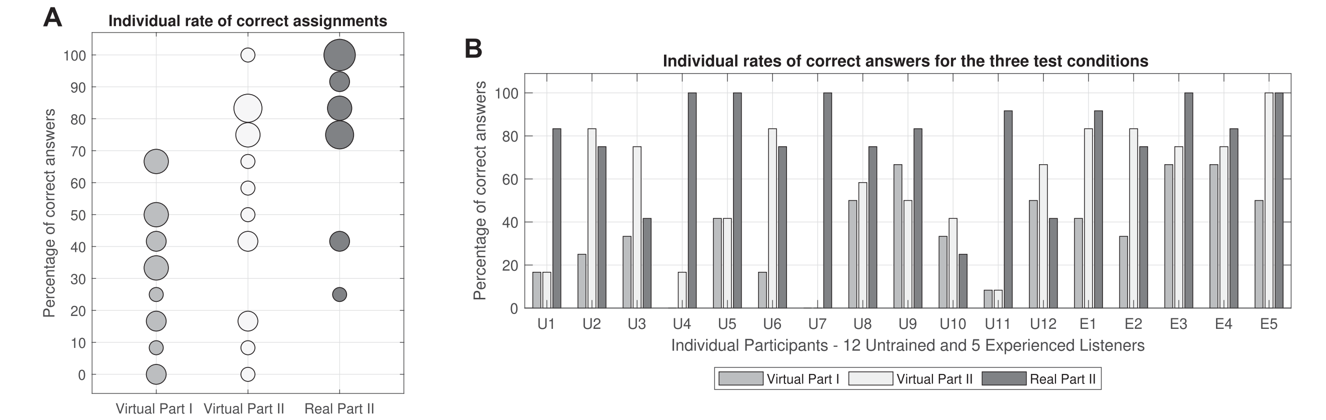

FIGURE 4 | (A) Individual rates of correct answers sorted by test condition. The size of the bubbles indicates how many subjects achieved this result. **(B)** Percentage of correct answers achieved by the 17 individual participants for the scenes with the virtual reproduction in part I and part II of the experiment and the real source in part II.

sensitivity. Only the overall percentage of correct answers was taken into account, assuming that the number of correct answers would be equally distributed over real and virtual scenes and considering equations developed for a 2AFC test design. A Yes/No paradigm differs from a 2AFC paradigm. In a Yes/No paradigm, the stimuli are presented and rated one by one. In contrast, the 2AFC paradigm as considered in the SDT offers Noise and Signal Stimuli (in our experiment, real and virtual) within one trial, in either randomized temporal or spatial order. Therefore, a 2AFC paradigm allows for direct comparison between both stimuli. Furthermore, the answer in each trial is correct or wrong for both stimuli at the same time. In contrast, in a Yes/No paradigm, distinguishing Hits and Correct Rejections can provide additional or more accurate information since they are not necessarily equal. **Figure 4** visualizes the individual percentage of correct answers of our experiment separated by real and virtual reproduction and shows that they are not equal. Therefore, we considered p_{Hit} and p_{FA} rather than only the percentage of correct answers. According to, e.g., Wickens (2001), p_{Hit} and p_{FA} can be calculated as follows:

$$p_{Hit} = \frac{\text{Number of Hits}}{\text{Total Number of Signal Presentations}}$$

$$p_{FA} = \frac{\text{Number of False Alarms}}{\text{Total Number of Noise Presentations}}$$

The sensitivity d' can be determined with the following equation:

$$d' = z(p_{Hit}) - z(p_{FA}) \tag{1}$$

This equation is a criterion-free estimation of the sensitivity. It can be used to determine the individual and the pooled sensitivity. For extreme values of p_{Hit} and p_{FA}, a correction according to Hautus (1995) was applied. This correction is integrated into the *dprime*-function in R, which we used for this analysis. Since the sample size per person is relatively small, both mean and pooled sensitivity will be estimated and compared. In addition, the decision criterion location c can be calculated. c indicates the distance of the decision criterion from the center between both distributions.

$$c = -\frac{1}{2}\left(z(p_{Hit}) + z(p_{FA})\right) \tag{2}$$

c is zero, if False Alarms and Misses occur with an equal percentage of the Noise and Signal samples. If c is below zero, there is a tendency towards the answer “virtual.” In contrast, a positive value indicates a tendency towards the answer “real.”

Another question is which value of d' indicates that the discriminability of the virtual reproduction is sufficiently small. Lindau and Weinzierl (2012) have determined such a minimum effect hypothesis under the assumption of non-biased participants and only considering the percentage of correct answers. For a group of subjects with considerable differences in individual bias, the determination becomes more challenging. Therefore, we additionally consider another interpretation of the data.

2.10.2 Analysis Based on the Paired *t*-Test

It is interesting to analyze the rate of acceptance as real. For the real source, this number is equivalent to the number of the correct answers. For virtual reproduction, it is the number of wrong answers. The auditory illusion can be considered plausible if the rate of acceptance for the virtual source does not vary significantly from that of the real sound source. In order to test for significant differences in the rates of acceptance between the real and the virtual test scenes, a paired *t*-test can be used. The *t*-test is suitable even for small sample sizes. The analysis considers the distribution of the individual rates of acceptance for both test conditions. The paired *t*-test assumes that the difference between both test conditions follows a normal distribution. This was tested and confirmed with a Shapiro–Wilk test, although it has to be noticed that testing for normal distribution can be inaccurate for small samples. The paired *t*-test checks whether the hypothesis that the two samples follow distributions with equal means can be rejected.

3 RESULTS

The auditory illusion of a loudspeaker reproducing sound is considered plausible if the listeners cannot identify it as virtual systematically. The realization of the position-dynamic binaural synthesis in this experiment does not contain any individualization of the BRIRs. Consequently, we expected that at least the experienced listeners would detect the virtual reproduction among the test scenes in this Yes/No paradigm. The study also aims at identifying available audible cues that can reveal the simulation. This is of interest for a targeted improvement of the system.

Furthermore, since considering a real reference in a perceptual evaluation comes with practical challenges and limitations, we want to know whether the availability of a real reference influences the estimated plausibility of the auditory illusion. For this reason, the experiment was conducted in two parts. The first evaluates the plausibility regarding the pure internal reference without considering real sound fields. The second part evaluates plausibility with the test approach proposed by Lindau and Weinzierl (2012) by including real versions of the simulated sound fields. Does the availability of the real sound field affect the plausibility?

3.1 Observations of the Informal Pretest

In the pre-test, three experts who did not participate in the subsequent main experiment listened to the real and the virtual version of the loudspeaker reproduction in a direct AB comparison for the various test cases listed in **Table 2**. The experts described freely which differences they perceived. It was interesting to notice that after a short episode of exploration, the experts moved to the closest position possible to the front of the active sound source. Once they arrived there, they focused on rotating their heads or turning themselves at that position. Sometimes, they reported a slight instability of the perceived location of the sound source during head rotation. Furthermore, when turning the back towards the sound source, differences between real and virtual reproduction were audible. The deviations were described as a change in distance perception, externalization, and relative sound level. For the binaural reproduction, the source was described to be in the head or sticking to the back of the head. However, with the real reproduction, the source in the back did not appear fully natural as well. The distance perception did also not match the expectations. In the AB comparison, the experts noticed minimal deviations in timbre, reverberance, and apparent source width in addition to the previously mentioned effects.

3.2 Overview and Individual Differences

In this experiment, each of the 17 subjects rated 36 test scenes. In total, these are 612 answers. 348 of these answers (56.9%) were correct. With 30 correct assignments out of 36 (83%), one of the trained listeners achieved the highest individual number of correct answers. The other experienced participants rated 23, 27, 28, and 29 scenes correctly in the course of the experiment. Two inexperienced listeners achieved the lowest individual rate of correct responses with 12 out of 36 (33%). These numbers indicate that identifying the virtual reproduction among the randomized test items was not an easy task. However, the numbers sum up different test cases that should be considered separately. The three main categories of test cases are "virtual sound source tested in part I of the experiment," "virtual sound source tested in part II of the experiment," and "real sound source tested in part II of the experiment." For each of these categories, each of the 17 participants rated 12 test scenes and achieved an individual number x_i of correct answers between 0 and 12.

Figure 4B illustrates the individual rates of correct answers each of the participants achieved in the three test conditions. The percentage of correct answers varies substantially among the participants. Furthermore, the distribution of the correct responses over the three test conditions is very different from person to person. **Figure 4A** basically shows the same numbers but sorted by test condition. The data for the separate conditions exhibit different trends. A paired t-test was conducted to test whether the sample of individually achieved rates of correct answers is part of distributions with equal means. According to the paired t-test, in part II of the experiment for the cases when the visible loudspeaker was actually reproducing the sound, the participants answered significantly ($t\ (16) = 2.24$, $p < 0.04$) more often correctly ($M = 9.47$, $SD = 2.74$) than for the test scenes with the virtual reproduction ($M = 6.76$, $SD = 3.68$). Furthermore, for the virtual scenes in part II of the experiment, the subjects answered significantly more often correctly, $t\ (16) = 3.50$, $p = 0.003$ ($M = 6.76$, $SD = 3.68$), than for the same test scenes in part I ($M = 4.24$, $SD = 2.63$).

3.3 Correct Identification of the Real Source and Its Limitations

First, it is of interest how often the participants identified the real sound source as real. Each of the 17 participants evaluated 12 test cases in which the sound source was real. This results in a total of 204 evaluations. Overall, in only 161 of the 204 assignments (78.9%), the participant chose real as the answer. This indicates that, at least for some of the listeners, the internal reference is not perfectly reliable. Probably, most participants have never paid attention to what it sounds like to walk towards or past a loudspeaker or turn around in front of it. Usually, listeners have a basic idea of what to expect but feel uncertain about the details. Additionally, the subjects had to listen to the real loudspeaker while wearing headphones. This is an uncommon listening situation for which most listeners might not have an adequate internal reference. Generally, real listening scenarios may exhibit details which the listener did not expect. Such elements may be mistaken as cues revealing the virtual sound source. Especially for listeners with no or little experience in the field of binaural technology, the task was challenging. The five experienced listeners correctly identified the real source in 12, 12, 11, 10, and 9 of the 12 test cases (on average 90.0%).

TABLE 4 | Estimated sensitivity d' and decision criterion c for both parts of the experiment. As expected, the sensitivity estimated for part I is considerably lower than for part II. For both parts, the decision criterion indicates a tendency towards the response "real." In part I, this tendency is even stronger than in part II.

Results for d' and c	Part II (tuned internal reference)	Part I (pure internal reference)
Mean sensitivity	1.05 (c = 0.37)	0.46 (c = 0.67)
Pooled sensitivity	0.96 (c = 0.32)	0.43 (c = 0.59)

Inexperienced listeners were correct in 74.3% of the cases. **Figure 4** visualizes the individual results. Three inexperienced listeners rated the real loudspeaker reproduction as real only in three or five of the 12 test cases. Especially, the person with the three correct identifications tended to assign virtual and real scenes vice versa.

3.4 Analysis of Part 2: Plausibility Evaluation With a Tuned Internal Reference

This part of the analysis focuses on part II of the experiment, where the plausibility was evaluated, including the real counterparts of the test scenes. This test design is in accordance with the method proposed by Lindau and Weinzierl (2012). They have determined the sensitivity index d' as an indicator of the discriminability between real and virtual versions of the scenes based on the Signal Detection Theory (SDT). We analyzed our results accordingly.

3.4.1 Estimating the Discriminability Based on Signal Detection Theory

The sensitivity index d' can be calculated with **Eq. (1)**. Due to the small sample size per person small, in addition to the common mean sensitivity, we determined the pooled sensitivity to compare both. The first column in **Table 4** shows the results for part II of the experiment. The mean sensitivity determined from the individual sensitivities of each participant differs only slightly from the pooled sensitivity, which was determined from the overall number of Hits and False Alarms. Both values are close to one and indicate good discriminability. The decision criterion c is determined with **Eq. (2)**. Due to the small sample size, also c was calculated as a mean of the individual response bias and as the pooled criterion overall. The difference between both values is minimal. The positive value shows that the location of the decision criterion is shifted towards the distribution of Hits. This indicates that in part II, the subjects had, on average, a tendency towards the response "real."

3.4.2 Analysis Based on the Paired *t*-Test

Figure 5 shows how often the participants picked the answer "real" in each of the conditions. This indicates the rate of acceptance as real. The paired t-test checks for the hypothesis that the two samples follow distributions with equal means. For the distributions of the individual acceptance rates as real, this hypothesis can be rejected, $t\ (16) = 4.18$, $p < 0.001$. This means, in part II, the acceptance of the virtual reproduction ($M = 5.23$, $SD = 3.68$) was significantly lower than that of the real reproduction ($M = 9.47$, $SD = 2.74$).

In addition to the results of the whole group of participants, **Figure 5** shows the separate results for experienced and inexperienced listeners. For both groups, the paired t-test separately still indicates significant differences between the acceptance of real and virtual reproduction, experienced $t\ (4) = 9.6$, $p < 0.001$ ($M_{real} = 10.80$, $SD_{real} = 1.30$ and $M_{virtII} = 2.0$, $SD_{virtII} = 1.23$) and inexperienced listeners $t\ (11) = 2.50$, $p < 0.05$ ($M_{real} = 8.92$, $SD_{real} = 3.03$ and $M_{virtII} = 6.58$, $SD_{virtII} = 3.53$). Although the numbers indicate that discriminability is quite good, the subjects found it hard to distinguish whether the loudspeaker was reproducing sound virtually or for real. They had the chance to take as much time as they needed to explore the scene and decide. An average duration of the exploration per scene of 70 s indicates that the decision was not taken right away. Providing a convincing auditory illusion of the given scenario that endures this high degree of interactivity and this long and intense exploration is a more critical test than a short one-time listening. Achieving plausibility with regard to a "tuned" internal reference is more challenging.

3.5 Analysis of Part I: Plausibility With Regard to the Pure Internal Reference

In **Figure 4B**, the first and second row of bubbles show the individual percentage of correct identifications of the virtual sound source in the first and the second part of the experiment. In part I, the case in which only the virtual sound source was presented, in 71 of the 204 test scenes (34.8%), the virtual sound source was identified correctly. In part II, the virtual sound source was presented alongside the real version in a randomized order. In this case, it was identified correctly in 115 of the 204 scene assignments (56.4%). The statistical analysis is again based on the two approaches, Signal Detection Theory and the paired t-test.

3.5.1 Analysis Based on Signal Detection Theory

In order to compare the evaluations of the virtual sound sources in part I and part II of the experiment in SDT, the sensitivities were calculated for both parts in relation to the evaluation of the real sound source conducted in part II. Thus, mean and pooled sensitivity were calculated again, this time with $pHit$ based on the rate of correct identifications of the virtual reproduction in part I instead of part II. **Table 4** provides an overview of the estimated sensitivities.

Again, mean and pooled sensitivity are very similar. As expected, the sensitivity estimated for part I is considerably lower than that for part II. For both parts of the experiment, the decision criterion indicates a tendency towards the response "real." In part I, this tendency is even stronger than in part II.

3.5.2 Analysis Based on the Paired *t*-Test

Considering the individual rates of acceptance as real, it is the question of whether there is a significant difference between the

FIGURE 5 | Percentage of test scenes which were rated as real by the 17 individual participants, as well as Inexperienced and Experienced Listeners separately for the scenes with the virtual reproduction in Part I and Part II of the experiment and the real scenes in Part II.

acceptance of the virtual scenes in part I and part II of the experiment. The results of the paired *t*-test indicate that over all subjects, the hypothesis of equal means can be rejected, $t\ (16) = 3.50$, $p < 0.005$. In part I, the acceptance of the virtual reproduction ($M = 7.76$, $SD = 2.63$) was significantly higher than in part II ($M = 5.24$, $SD = 3.68$). This holds for both experienced ($t\ (4) = 3.28$, $p < 0.04$) ($M_{virtI} = 5.80$, $SD_{virtI} = 1.79$ and $M_{virtII} = 2.0$, $SD_{virtII} = 1.22$) and inexperienced ($t\ (11) = 2.25$, $p < 0.05$) ($M_{virtI} = 8.58$, $SD_{virtI} = 2.54$ and $M_{virtII} = 6.58$, $SD_{virtII} = 3.53$) listeners. This result is not surprising. An influence of real scenes among the test items was expected.

In addition, it is of interest to compare the results of part I to those of the real scenes. For a significance level $\alpha = 0.05$, the hypothesis of equal means cannot be rejected, $t\ (16) = 1.94$, $p = 0.07$. Thus, the acceptance of the virtual reproduction in part I of the experiment ($M = 7.76$, $SD = 2.63$) is not significantly different from the acceptance of the real scenes in part II ($M = 9.47$, $SD = 2.74$). This is an exciting observation. Taking only the experienced listeners into account, the paired *t*-test indicates that the means of the acceptance of virtual scenes in part I ($M = 5.80$, $SD = 1.79$) and real scenes in part II ($M = 10.80$, $SD = 1.30$) differ significantly, $t\ (4) = 4.23$, $p = 0.01$. The bubble chart in **Figure 5** visualizes the individual acceptance rates for experienced listeners. The rates are visually quite well separated for the three test conditions.

For inexperienced listeners, the paired *t*-test does not reject the hypothesis of equal means at all, $t\ (11) = 0.34$, $p > 0.7$ ($M_{virtI} = 8.58$, $SD_{virtI} = 2.54$ and $M_{real} = 6.58$, $SD_{real} = 3.53$). This means that the created spatial auditory illusion is convincing enough that inexperienced listeners do not notice it is an illusion when relying purely on their internal references. This observation is essential for future studies with the goal of evaluating plausibility.

3.6 Cues Used for Detection of the Virtual Reproduction

Figure 6 provides a summary of the audible cues mentioned by the participants in the interview after the test. This overview does not consider the relation to the individual detection rates but represents all answers given by the subjects.

Twelve of the 17 subjects reported that the sound source behaved unnaturally when they turned their backs towards it. The source appeared closer, sometimes even in the head, and varied in loudness. This observation is in line with the reports by the trained listeners in the pre-listening session.

Nine participants reported an unnatural experience of head rotation. The source position appeared slightly unstable. The effect increased with the speed of rotation. Seven of the subjects stated that this was the main cue they used to identify the binaural auralization. This observation is also in line with the effects reported by the experts in the pre-test.

In addition to these two major cues, some participants reported that they perceived the sound source in the head before they started to move. Some listeners mentioned that the sound level changed in a way they did not expect. Few people stated that they perceived differences in timbre, apparent source width, and localizability.

3.7 Source Position, Type of Signal, and Sound Level

Figure 7 provides an overview of the rates of correct answers with respect to the source position, the type of signal, and the sound level. Only part II of the experiment is considered for this analysis.

The first graph visualizes the subjects' individual rates of correct answers within the test. For each source position and each sound level condition, the total number of test cases per person was twelve, six real and six virtual. According to the paired *t*-test, the percentage of correct answers for the source in the front ($M = 7.94$, $SD = 2.38$) does not differ significantly, $t\ (16) = 1.0$, $p > 0.3$, from the percentage of correct answers for the lateral source position ($M = 8.29$, $SD = 2.02$). The percentage of correct answers for the 0 dB sound level ($M = 8.24$, $SD = 2.22$) was not significantly different, $t\ (16) = 0.57$, $p > 0.3$, from that for the −6 dB ($M = 8.0$, $SD = 2.29$).

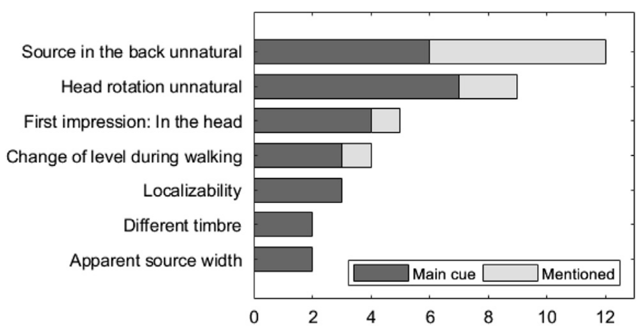

FIGURE 6 | Overview of audible cues reported to be used by the participants to discriminate the binaural simulation from the real sound field.

For each type of signal, each participant rated eight scenes in part II, four virtual and four real. The individual rates of correct answers for speech (M = 5.29, SD = 1.61), music (M = 5.47, SD = 1.70), or snare (M = 5.47, SD = 1.94) were not significantly different from each other, t (16) < 0.5, p > 0.6, for all three combinations. In summary, the position of the sound source, the type of signal, and the sound level did not significantly influence the percentage of correct answers.

For the main question in this experiment, the percentage of correct answers gives only limited insight. So instead, it is of interest to analyze the acceptance as real. A separate analysis of the individual amount of correct answers for virtual and real scenes for each condition was not feasible. This is because the sample size per person is already quite small for all of them together. However, a pooled inspection is possible. **Figure 7** visualizes the pooled rate of scenes accepted as real per condition separated by virtual and real reproduction for the whole pool of participants. Again, only the results of part II of the experiment were taken into account. In addition to the bars indicating the percentage correct for each condition, the confidence intervals proposed by Clopper and Pearson (1934) are shown. The virtual sources were accepted as real significantly less often than the real source for each of the conditions. There is an overlap of the CIs for the correct identification of the real scenes (SDT: Correct Rejections), and also, the percentage of virtual scenes accepted as real (SDT: Misses) does not vary significantly with source position, type of signal, or sound level.

In summary, neither source position nor the level or type of signal had a significant impact on the plausibility. This is especially interesting regarding the source position, considering that with the source position, the listener's motion relative to the loudspeaker was different. For the frontal sound source, the subjects could walk towards and away from it. For the position right of the translation line, the participants could walk past the front of the loudspeaker. The directivity of the sound source has a substantial impact on the progress of the direct sound. These differences between the test conditions did not exhibit different quality in terms of plausibility as the agreement with the tuned internal reference.

4 DISCUSSION

In this experiment, the plausibility of an auditory AR illusion created over headphones for a position-dynamic exploration by the listener was evaluated with regard to the pure internal reference on the one hand and with regard to an internal reference that was tuned by including the real counterpart of the test scenes on the other hand.

4.1 Plausibility of Position-Dynamic AAR Realization

When the real test scenes were included as hidden references, experienced listeners could identify binaural auralization quite confidently and inexperienced listeners did not predominantly accept the virtual reproduction as real anymore as in part I.

One of the main cues to identify the auralization was the audible difference in case the listener turned his back towards the source. Distance perception, externalization, and timbre were affected. All the previous studies did not document such an effect. Brinkmann et al. (2017) have tested the authenticity for source directions of 0° and 90°, allowing a head rotation of ±34°. The study was conducted with the extra-aural headphones. Lindau and Weinzierl have worked with STAX headphones and allowed a head rotation of ±80°. Pike et al. (2014) have also used STAX and provided a system capable of a full 360° reproduction, but instructed their participants to move only their heads but keep their torso still. The case of the source in the back has not received any attention so far. This means that our study is also the first we know to investigate plausibility with regard to the tuned internal reference for dynamic binaural synthesis with "true 360°." It is hard to tell whether the observed effect in the back is unique in the system used for this study or whether it is a general phenomenon. In the previous studies, AKG K1000 headphones were not used. Satongar et al. (2015) have shown that the passive influence of headphones can cause spectral distortions, affect the effective interaural time difference, and reduce localization accuracy. However, their study did not consider the AKG K1000. Measurements of the physical effect of AKG K1000 headphones by Pörschmann et al. (2019) and Schneiderwind et al. (2021) indicate that these might contribute to such audible effects.

Another cue was the slight instability of the source position during quick head rotation. Similar observations were reported in an earlier study by Lindau and Weinzierl Lindau et al. (2007) testing an early-stage system, as well as by Pike et al. (2014). This audible effect could be due to non-individualized ITDs or a non-optimal delay in the motion-related updating of the BRIR filters. These aspects have to be improved to achieve an authentic or plausible (with regard to tuned internal reference) reproduction.

Five subjects mentioned that they localized the sound source in the head before starting to move. They assigned this experience to the binaural simulation. However, in-the-head localization can occur in real sound fields as well (Plenge, 1972). It is questionable whether this is a reliable cue for the identification of virtual sound

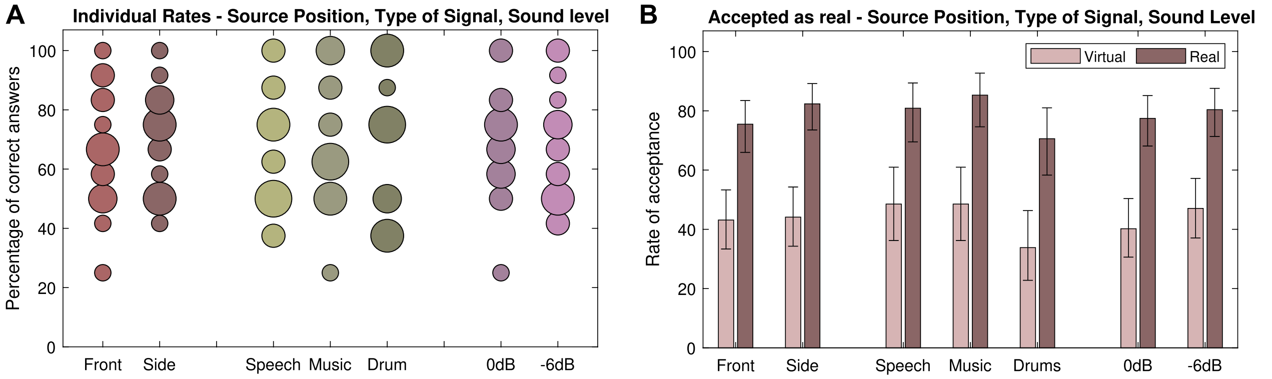

FIGURE 7 | (A) The individual rates of correct assignments were not significantly influenced by the position of the sound source, the type of source signal or the sound level - **(B)** No significant influence of source position, type of signal and sound level on the acceptance of the real and the virtual sound source in Part II could be observed in this experiment.

sources. Still, it may occur more often or more pronounced in a binaural reproduction.

Four participants stated that for them, the change of level during walking was a helpful cue. They reported that the level would change not enough or too much over certain sections of the translation line. These effects were also reported in previous experiments on the plausibility of an approaching motion Neidhardt et al. (2018); Kamandi (2019). Therefore, several untrained listeners were surprised about the progress of the sound level in the measured scenario and rated manipulated version of the scene as more plausible because the level change was closer to what they expected. This may also be a case of an inaccurate or wrong internal reference. In fact, also in the present experiment, this cue was only reported by untrained listeners.

Three participants reported a confusing localization that includes increased elevation (higher than the visual source) and reduced sharpness in the image of the sound sources. An increased elevation in the localization is a common artifact in the binaural simulation with non-individual BRIRs. This is likely to be a reliable cue revealing the simulation for some people. An increased blurriness might result from reproduction with generic BRIRs as well.

Furthermore, two participants perceived differences in the timbre and stated that the simulation has less strength in the low frequencies. The stimuli were limited to a frequency range between 150 Hz and 16 kHz for both reproduction methods. Deviating timbre might be caused by the non-individual BRIRs and a non-individual headphone compensation.

Two people reported an increased apparent source width. This usually occurs with an increase of reverberant energy. However, these reports may be connected to the reduced sharpness of the sound image when listening to a real sound source while wearing headphones.

This experiment was the first to consider position-dynamic binaural synthesis and their corresponding real version of the sound field in a test scenario with interactive self-translation of the listener. Furthermore, this study was the first to consider a true 360° experience when studying the discriminability of the auditory illusion from its real version.

The majority of the cues reported as helpful for identifying the virtual version were not related to translation. Four of the untrained subjects mentioned that the sound level would exhibit unexpected progress during walking. Similar statements were given in a previous experiment by (Kamandi, 2019) for the measured scene by participants who rated another artificial scene with a considerably greater change of the level as plausible. This judgment may be the result of an inaccurate or wrong internal reference. 13 of the 17 subjects in the present experiment did not mention any translation-related cues at all. Thus, the present realization of the translation did not cause substantial effects revealing the binaural auralization. However, without the additional freedom of motion in this test scenario, the observation regarding the unnatural impression of the sources in the back may not have been possible. In addition, it is interesting noticing that no significant differences between the cases of walking past and towards/away from the loudspeaker were observed.

4.2 Influence of the Availability of the Real Version: Pure *Versus* Tuned Internal Reference

Creating a test design investigating the influence of the availability of real versions of the sound source on the estimated plausibility is not straightforward. It has to be taken into account that the test without the real reproduction always had to be conducted first and without any training. Especially for inexperienced listeners, it is likely that it takes a while to identify helpful cues and establish strategies for efficient exploration. Such effects could not be eliminated with the given test design. Then again, it is possible that the identification of helpful cues revealing the virtual scene is easier when a real scene is presented in between. For the progress of the share of correct answers over the trials in the tested order, a regression analysis was conducted. This analysis is independent of the actual test condition. Both parts of the experiment were analyzed separately. The hypothesis that the regression coefficient is zero could not be rejected ($p > 0.6$ in both cases). This indicates a flat "learning curve" with no trend or evident increase in the number of correct answers in the course of the experiment. Consequently, it is reasonable to neglect the effects of training or getting used to the task for conclusions based on the submitted answers.

Another influence might be an expectation of the participants that real and virtual test scenes may be equally distributed in the test sample or at least a certain minimum amount of both options is included. This might have an effect if, in part I, subjects are not sure of the answer and become irritated by having the impression of repeatedly listening to the virtual version. In these cases, subjects might answer "real," although they actually tend to answer "virtual." However, this is only an issue if a subject cannot confidently identify virtual reproduction. In contrast, at least several of the inexperienced listeners answered with real very often. Apparently, they did not mind giving the same reply repeatedly.

To minimize this issue in part I, after 12 virtual scenes, 12 real scenes should be tested in addition. Then, part II with the same scenes in randomized order could follow. In that setup, the number of correct answers for the 12 real scenes in a row would be affected by the same psychological bias. The percentage of correct answers and thus the rate of acceptance would be reduced. Comparing the results of this part to the purely virtual part in terms of the paired t-test or calculating the sensitivity index would be less critical than comparing it to the results of the real scenes in the part with the randomized order. We decided not to include such a part in the experiment because the test was quite long already. Instead, we chose to use a more critical evaluation by comparing the results of part I to the real scenes in part II. We assume that the main findings of this experiment are not affected by this decision.

The results of this experiment suggest that including the real version of the scenes affects the listener's capability of identifying the simulation. The test design with randomized order of different signals, source positions, and sound levels minimized the options for a direct comparison between a virtual scene and its real counterpart. Thus, we can conclude that the test design influences the internal reference, which is fundamental for evaluating plausibility.

The fact that including the real version affects the estimated plausibility and reduces the acceptance of the virtual imitation is not surprising. It is known from other test methods that the choice of test items influences the test results for the single items and that including a (hidden) reference representing the best possible quality facilitates critical testing as discussed, for example, by Zielinski et al. (2008). The observations indicate that in the future, discriminating between different kinds of plausibility may be of interest. On the one hand, the plausibility that measures the agreement with the listener's pure internal reference will be of interest, e.g., in the case of fictive scenes. On the other hand, the plausibility that measures the agreement with an internal reference tuned by listening to a real version of the scene will allow for a more critical evaluation.

In augmented acoustic reality, the real environment is always present and will provide a kind of reference for a virtual acoustic element. For evaluating its quality, it is important to consider the influence of the elements and properties of the real acoustic environment. Authenticity is evaluated in a direct comparison of a virtual and a real scene and is therefore even more sensitive.

4.3 How Should the Plausibility of Auditory AR Be Evaluated in the Future?

This study considers an AAR scene, which contains one primary sound source besides the common quiet background sound in everyday environments like the chosen seminar room. The participants experienced the room with its acoustic behavior when they entered the room, walked to the test setup, talked to the test conductor, and got the introduction. This is likely to cause certain expectations towards how the reproduction of the loudspeaker standing in the room should sound. However, more complex scenes which contain a variety of real and virtual sound sources are more interesting and more common for application scenarios of AAR. There is usually no option in such scenarios to listen to exactly the real version of the virtual sound element at exactly the same position. Instead, the real sound sources of the actual acoustic environment are available among the virtual contents and serve as an external reference to some extent. Wirler et al. (2020) have already shown that the scene complexity affects the plausibility evaluations. The results of our study suggest that an available real equivalent to the virtual sound object will have a tuning effect on the internal reference. Further studies are necessary to improve the understanding of a listener's internal reference and its interrelation with different types of external reference. This is especially interesting in the case of fictional contents in terms of how their perception and acceptance are influenced by the other real and virtual elements of the given scenario.

Evaluating plausibility with regard to the pure internal reference has the advantage that a consideration of the headphones in the BRIR measurement is not required. In this experiment, headphones had to be taken into account to focus the investigation on the test method and avoid changing more than the primary variable among the test conditions. However, apart from the significant differences between both test methods, we observed that the main cue for identifying the virtual reproduction among the real scenes was probably caused by the shadowing effect of the headphones. This raises the question, whether the significant differences in plausibility hold if the evaluation with respect to the pure internal reference was conducted with BRIRs neglecting the occlusion effect. With regard to the desired ecological validity of test methods in general, both methods are of equal interest. For AR, the listener will always have to wear some sort of listening device. Despite all attempts to create a transparent headphone experience, perfect transparency has not been achieved yet. Then again, the overall goal is to create auditory illusions that appear as in the real world without the slight influences of any headphones.

4.4 Summary

The experiment presented in this article was conducted to evaluate the plausibility of walk-through scenarios with position-dynamic binaural synthesis using a state-of-the-art system. The realization is based on BRIR filters measured with a Kemar head and torso simulator wearing AKG K1000 headphones in the room and at the positions where the psychoacoustic experiment took place. The subjects could see two loudspeakers in the room, and in each scene, one of them reproduced sound either virtually or in reality. The subjects could either walk past the sound source or towards and away from it in different test cases. Head rotation and self-rotation were possible at all times. The subjects had to determine whether they heard the real reproduction or its binaural simulation in each trial. Dry male speech, a snare drum sample, and music in terms of a pop song were investigated. The experiment was divided into two parts. In part I, the plausibility was evaluated with regard to the subject's pure internal reference without the option to listen to a corresponding real version of the simulated sound fields. In part II, the approach of discriminating the binaural auralization from the corresponding real sound fields, as proposed by Lindau and Weinzierl (2012), was applied to binaural walk-through scenarios with a true 360° experience for the first time. Including real sound scenes as test items is accompanied by some challenges and limitations. On the one hand, the method can only consider the real scene as it is perceived through the used headphones or hearables. On the other hand, these effects have to be considered in the creation of the auditory illusions, for example, by measuring an extra set of BRIR measurements, including the hearing device of interest. Moreover, the method can only consider contents where a corresponding real version is available. In three earlier studies, the given system has repeatedly been rated as plausible in an evaluation without any real scene. If no real scene is included, it is not necessary to take the occlusion or shadowing effects of the headphones into account in the creation of the virtual content. Thus, there is no optimal evaluation method. In addition to the previous experiments, the present study evaluates the plausibility in a Yes/No paradigm with and without including the real versions of the simulated scenes as hidden references.

With the given AAR system, the inexperienced listeners accepted the virtual version as real in most cases in part I when the real scenes were not available. Even the experienced listeners could not confidently identify the presentation as a simulation in this case. In contrast, in part II, when the real versions were available in the test, experienced listeners could detect the simulation quite confidently while inexperienced listeners at least increasingly doubted the realness in the case of the virtual version. Source position, type of walking motion relative to the source, type of the source signal, and its sound level did not significantly influence the observations. Two primary cues revealed the virtual reproduction. In the listener's back, the sound source exhibited an unnatural appearance, which was caused by the presence of the headphones. In addition, the participants reported slight instabilities of the sound source during head rotation, which were probably caused by the lack of individualization and maybe a non-optimal system latency.

4.5 Conclusion

The results of the presented study indicate that the system under test is capable of inducing a plausible illusion for inexperienced listeners. However, the system fails to deliver a plausible illusion for experienced listeners in general and for all listeners if they had the chance to listen to the real counterpart of the sound field. The primary cues affecting plausibility are not caused by the increased freedom of motion of this AAR setup but rather introduced by the presence of the headphones and the lack of individualization. As expected, the results show that the availability of a real counterpart tunes the internal reference and leads to a more critical evaluation of plausibility. On the one hand, this suggests that the presence of similar real sound objects in an AR scenario may also affect the plausibility of virtual content. On the other hand, this evaluation method demands considering the occlusion effect of the headphones in the synthesis of the virtual content. This reduces the overall quality of the AR reproduction and limits the ecological validity of this test approach. However, the fact that perfectly transparent headphones are not available remains a challenge for realizing AR systems. Especially for motion in 6DOF, the knowledge about this influence on the perception of real sound sources is still surprisingly low. Under these test conditions and compared to these effects, potential imperfections of the position-dynamic binaural synthesis used in the system under test did not appear critical for the plausibility of the AAR realization.

DATA AVAILABILITY STATEMENT

The dataset of measured BRIRs used in this study can be found at Zenodo *https://zenodo.org/record/3457782*. The pyBinSim software Neidhardt et al. (2017) for dynamic binaural auralization is available at https://github.com/pyBinSim or https://pypi.org/project/pybinsim/.

ETHICS STATEMENT

Ethical review and approval were not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

As AN's research focuses on the perceptual issues of listener translation motion and an efficient realization of dynamic binaural synthesis for an exploration in six degrees of freedom, she initiated this study and formulated the main research question. She was the supervisor of the master thesis in which this study was conducted and she also wrote the text in this article. AMZ realized the measurements and conducted the experiment as main task in her master thesis. In addition, she contributed considerably to the review of previous experiments and in taking decisions for many details of the test design. The final version of her thesis inspired several of the sections and the discussion in this article.

FUNDING

The study was funded by DFG (Deutsche Forschungsgemeinschaft, BR 1333/18-1).

ACKNOWLEDGMENTS

The authors would like to thank all volunteers for participating in this experiment. Special thanks go to Christoph Pörschmann and Christian Schneiderwind for the very helpful discussions and proofreading.

REFERENCES

Azuma, R. T. (1997). A Survey of Augmented Reality. *Presence: Teleoperators Virtual Environ.* 6 (4), 355–385. doi:10.1162/pres.1997.6.4.355

Blauert, J. (1997). "Spatial Hearing," in *The Psychophysics of Human Sound Localization* (Cambridge, MA, USA: MIT Press). revised edition.

Brandenburg, K., Klein, F., Neidhardt, A., Sloma, U., and Werner, S. (2020). "Creating Auditory Illusions with Binaural Technology," in *The Technology of Binaural Understanding*. Editors J. Braasch and J. Blauert (Cham, Switzerland: Springer Nature Switzerland AG), 623–663. doi:10.1007/978-3-030-00386-9_21

Brinkmann, F., Lindau, A., and Weinzierl, S. (2017). On the Authenticity of Individual Dynamic Binaural Synthesis. *J. Acoust. Soc. Am.* 142, 1784–1795. doi:10.1121/1.5005606

Clopper, C. J., and Pearson, E. S. (1934). The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial. *Biometrika* 26, 404–413. doi:10.2307/233198610.1093/biomet/26.4.404

Enge, K., Frank, M., and Höldrich, R. (2020). "Listening experiment on the Plausibility of Acoustic Modeling in Virtual Reality," in In 46th Annual Meeting on Acoustics (DAGA), Hannover, Germany, March 16–19, 2020.

Erbes, V., Schultz, F., Lindau, A., and Weinzierl, S. (2012). "An Extraaural Headphone System for Optimized Binaural Reproduction," in In 38th Annual Meeting on Acoustics (DAGA), Darmstadt, Germany, March 19–22, 2012.

Garí, S., Brimijoin, W., Hassanger, H., and Robinson, P. (2019). "Flexible Binaural Resynthesis of Room Impulse Responses for Augmented Reality Research," in In EAA Conference on Spatial Audio Signal Processing, Paris, France, September 6–7, 2019.

Hartmann, W. M., and Wittenberg, A. (1996). On the Externalization of Sound Images. *J. Acoust. Soc. Am.* 99 (6), 3678–3688. doi:10.1121/1.414965

Hautus, M. J. (1995). Corrections for Extreme Proportions and Their Biasing Effects on Estimated Values Ofd′. *Behav. Res. Methods Instr. Comput.* 27, 46–51. doi:10.3758/bf03203619

Heller, F., Jevanesan, J., Dietrich, P., and Borchers, J. (2016). "Where Are We?," in Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services, Florence, Italy, September 6–9, 2016. doi:10.1145/2935334.2935365

Herzog, M., Francis, G., and Clarke, A. (2019). *Understanding Statistics and Experimental Design - How to Not Lie with Statistics*. Germany: Springer Open Books. Learning Materials in Biosciences.

Hofer, M., Hartmann, T., Ratan, R., Eden, A., and Hahn, L. (2020). The Role of Plausibility in the Experience of Spatial Presence in Virtual Environments. *Front. VR* 1, 1–9. doi:10.3389/frvir.2020.00002

Jot, J., and Lee, K. (2016). "Augmented Reality Headphone Environment Rendering," in 2016 Aes International Conference On Audio For Virtual And Augmented Reality, Los Angeles, CA, USA, September 30–October 1, 2016.

Kamandi, S. (2019). Perception of Simplifications of the Room Acoustics in a Dynamic Binaural Synthesis for Listener Translation. Master's thesis. Germany: Technische Universität Ilmenau.

Kuhn-Rahloff, C. (2011). Prozesse der Plausibilitätsbeurteilung am Beispiel ausgewählter elektroakustischer Wiedergabesituationen. Ph.D. thesis. TU Berlin, Germany.

Langendijk, E. H. A., and Bronkhorst, A. W. (2000). Fidelity of Three-Dimensional-Sound Reproduction Using a Virtual Auditory Display. *J. Acoust. Soc. Am.* 107 (1), 528–537. doi:10.1121/1.428321

Latoschick, M., and Wienrich, C. (2021). Coherence and Plausibility, Not Presence?! Pivotal Conditions for XR Experiences and Effects, a Novel Model. *arXiv* [Epub ahead of print].

Lindau, A. (2009). "The Perception of System Latency of Dynamic Binaural Synthesis," in NAG/DAGA International Conference on Acoustics, Rotterdam, The Netherlands, March 23–26, 2009.

Lindau, A., Hohn, T., and Weinzierl, S. (2007). "Binaural Resynthesis for Comparative Studies of Acoustical Environments," in In 122nd International AES Convention, Vienna, Austria, May 5–8, 2007.

Lindau, A., and Weinzierl, S. (2012). Assessing the Plausibility of Virtual Acoustic Environments. *Acta Acustica united with Acustica* 98, 804–810. doi:10.3389/fnins.2013.1234510.3813/aaa.918562

Macmillan, N., and Creelman, C. (2004). *Detection Theory - a User's Guide*. New Jersey: Psychology Press.

Maseiro, B. (2012). Individualized Binaural Technology. Measurement, Equalization and Perceptual Evaluation. Ph.D. thesis. Germany: RWTH Aachen.

Moore, A., Tew, A., and Nicol, R. (2010). An Initial Validation of Individualized Crosstalk Cancellation Filters for Binaural Perceptual Experiments. *J. Audio Eng. Soc.* 58, 36–45. doi:10.1121/1.5005606

Nagele, A. N., Bauer, V., Healey, P. G. T., Reiss, J. D., Cooke, H., and Cowlishaw, T. (2021). Interactive Audio Augmented Reality in Participatory Performance. *Front. Virtual Real.* doi:10.3389/frvir.2020.610320

Neidhardt, A. (2019). "BRIRs for Position-Dynamic Binaural Synthesis Measured in Two Rooms," in In 5th International Conference on Spatial Audio (ICSA), Ilmenau, Germany. Data set.

Neidhardt, A., Klein, F., Knoop, N., and Köllmer, T. (2017). "Flexible python Tool for Dynamic Binaural Synthesis Applications," in In 142nd International AES Convention, Berlin, Germany, May 20–23, 2017.

Neidhardt, A., and Knoop, N. (2017). "Binaural Walk-Through Scenarios with Actual Self-Walking Using an HTC Vive," in In 43rd Annual Conference on Acoustics, Kiel, Germany, March 6–9, 2017.

Neidhardt, A., Tommy, A., and Pereppadan, A. (2018). "Plausibility of an Interactive Approaching Motion towards a Virtual Sound Source Based on Simplified BRIR Sets," in In 144h International AES Convention, Milan, Italy, May 23–26, 2018.

Oberem, J., Maseiro, B., and Fels, J. (2016). Experiments on Authenticity and Plausibility of Binaural Reproduction Ia Headphones Employing Different Recording Methods. *Appl. Acoust.* 114, 71–78. doi:10.1016/j.apacoust.2016.07.009

Pike, C., Melchior, F., and Tew, T. (2014). "Assessing the Plausibility of Non-individualized Dynamic Binaural Synthesis in a Small Room," in In 55th International AES Conference, Helsinki, Finland, August 27–29, 2014.

Plenge, G. (1972). On the Problem of "in Head Localization". *Acta Acustica united with Acustica* 26, 241–252.

Pörschmann, C., Arend, J., and Gillioz, R. (2019). "How Wearing Headgear Affects Measured Head-Related Transfer Functions," in In 1st EAA Spatial Audio Signal Processing Symposium, Paris, France, September 6–7, 2019.

Remaggi, L., Kim, H., Neidhardt, A., Hilton, A., and Jackson, P. (2019). "Perceived Quality and Spatial Impression of Room Reverberation in VR Reproduction from Measured Images and Acoustics, September 9–13, 2019.

Russell, S., Dublon, G., and Paradiso, J. (2016). "Hearthere - Networked Sensory Prosthetics through Auditory Augmented Reality," in 18th International

Conference on Human-Computer Interaction with Mobile Devices and Services, Florence, Italy, September 6–9, 2016. doi:10.1145/2875194.2875247

Satongar, D., Pike, C., Lam, Y., and Tew, A. (2015). The Influence of Headphones on the Localization of External Loudspeaker. *J. Audio Eng. Soc.* 63 (10), 799–810. doi:10.17743/jaes.2015.0072

Schärer, Z., and Lindau, A. (2009). "Evaluation of Equalization Methods for Binaural Signals," in In 126th International AES Convention, Munich, Germany, May 7–10, 2009.

Schneiderwind, C., Neidhardt, A., and Meyer, D. (2021). "Comparing the Effect of Different Open Headphone Models on the Perception of a Real Sound Source," in In 150th International AES Convention, May 25–29, 2021. Online.

Sicaru, I., Ciocianu, C., and Boiangiu, C.-A. (2018). A Survey on Augmented Reality. *J. Inf. Syst. Operations Manag.* 11 (2), 355.

Skarbez, R., Jr., Brooks, F., and Whitton, M. (2017). A Survey of Presence and Related Concepts. *ACM Comput. Surv.* 50 (6), 1–39. doi:10.1145/3134301

Slater, M. (2018). Immersion and the Illusion of Presence in Virtual Reality. *Br. J. Psychol.* 109, 431. doi:10.1111/bjop.12305

Slater, M. (2009). Place Illusion and Plausibility Can lead to Realistic Behaviour in Immersive Virtual Environments. *Phil. Trans. R. Soc. B* 364, 3549–3557. doi:10.1098/rstb.2009.0138

Stecker, G., Moore, T., Folkerts, M., Zotkin, D., and Duraiswami, R. (2018). "Towards Objective Measures of Auditory Co-immersion in Virtual and Augmented Reality," in AES International Conference on Audio for Virtual and Augmented Reality, Redmond, WA, USA, August 20–22, 2018.

Wefers, F., and Vorländer, M. (2018). Flexible Data Structures for Dynamic Virtual Auditory Scenes. *Virtual Reality* 22, 281. doi:10.1007/s10055-018-0332-9

Wickens, T. (2001). *Elementary Signal Detection Theory*. Germany, USA: Oxford University Press.

Wirler, S., Meyer-Kahlen, N., and Schlecht, S. (2020). "Towards Transfer-Plausibility for Evaluating Mixed Reality Audio in Complex Scenes," in AES International Conference on Audio for Virtual and Augmented Reality, August 17–19, 2020.

Zielinski, S., Rumsey, F., and Bech, S. (2008). On Some Biases Encountered in Modern Audio Quality Listening Tests - a Review. *J. Audio Eng. Soc.* 56 (6), 427–451.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Virtual Reality

ORIGINAL RESEARCH
published: 07 September 2021
doi: 10.3389/frvir.2021.656473

Check for updates

Social Presence Outside the Augmented Reality Field of View

Mark Roman Miller[1]* *and Jeremy N. Bailenson*[2]

[1]*Department of Computer Science, Stanford University, Stanford, CA, United States,* [2]*Department of Communication, Stanford University, Stanford, CA, United States*

Augmented reality headsets in use today have a large area in which the real world can be seen, but virtual content cannot be displayed. Users perceptions of content in this area is not well understood. This work studies participants perception of a virtual character in this area by grounding this question in relevant theories of perception and performing a study using both behavioral and self-report measures. We find that virtual characters within the augmented periphery receive lower social presence scores, but we do notfind a difference in task performance. These findings inform application design and encourage future work in theories of AR perception and perception of virtual humans.

Keywords: augmented reality, field-of-view, unaugmented periphery, social presence, social facilitation, social inhibition

Edited by:
Mary C. Whitton,
University of North Carolina at Chapel Hill, United States

Reviewed by:
Henrique Galvan Debarba,
IT University of Copenhagen, Denmark
Jessica Good,
Davidson College, United States

***Correspondence:**
Mark Roman Miller
mrmillr@stanford.edu

Specialty section:
This article was submitted to Virtual Reality and Human Behaviour, a section of the journal Frontiers in Virtual Reality

Received: *20 January 2021*
Accepted: *23 August 2021*
Published: *07 September 2021*

Citation:
Miller MR and Bailenson JN (2021) Social Presence Outside the Augmented Reality Field of View. Front. Virtual Real. 2:656473. doi: 10.3389/frvir.2021.656473

INTRODUCTION

One defining aspect of augmented reality (AR) is the integration of real and virtual content (Azuma, 1997). This integration is what separates AR from virtual reality and enables its unique applications. The development of AR headsets progresses towards a vision of computationally-mediated stimuli produced in fidelity high enough to be indistinguishable from reality.

However, this vision has not yet come to fruition. The limitation with which this work concerns itself is the field-of-view (FOV). In all headsets in use today, there are regions of the visual field in which real objects are visible but virtual objects are not. This region has the technical name *unaugmented periphery* (Lee et al., 2018), defined as the area within the real-world FOV that is not included in the virtual FOV. **Figure 1** illustrates these ranges from both a bird's-eye and a first-person view.

The narrow FOV of current headsets hampers interaction with virtual content. Anecdotally, we have observed first-time users become surprised and occasionally frustrated at the narrow FOV. To illustrate the effects of a narrow FOV, we provide a qualitative illustration of headset use in an example environment (**Figure 2**) and a quantitative calculation of the size of the unaugmented periphery in the Microsoft HoloLens.

To calculate the unagumented periphery, the real-world FOV must be compared with the virtual FOV. Virtual FOV has a straightforward measurement because it is often reported as a technical specification. The HoloLens virtual FOV is about 30° horizontally by 18° vertically (Kreylos, 2015).

The real-world FOV must be estimated based on the device and the size of the human visual field. An estimate for the real-world FOV for this device is approximately 180° horizontally by 100° vertically. This estimate comes in two parts, horizontal and vertical. The vertical field of view is limited slightly by the headset – an object at the very top edge of the human visual field (about 50° above horizontal) would be occluded by the headset's brim. A fair estimate of the maximum vertical angle visible through the HoloLens is about 30° above horizontal. Combined with the fact that minimum vertical angle is about 70° below horizontal (Jones et al., 2013), a fair estimate is approximately 100° vertical.

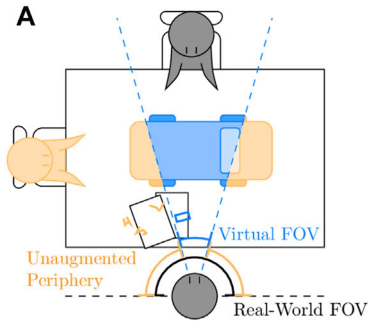

FIGURE 1 | Diagram of virtual and real-world fields of view. Virtual objects are drawn in blue and orange. Panel **(A)** shows a bird's-eye view of a team collaborating over a table. On the table is a virtual object. The virtual FOV (blue dotted lines) is smaller than the real-world FOV (short grey lines), leading to large sections of virtual content in the unaugmented periphery (orange). Panel **(B)** shows a first-person view, denoting the virtual FOV (blue) and unaugmented periphery (orange). The grey character is visible, being physically present, while the orange character is not visible, being in the unaugmented periphery.

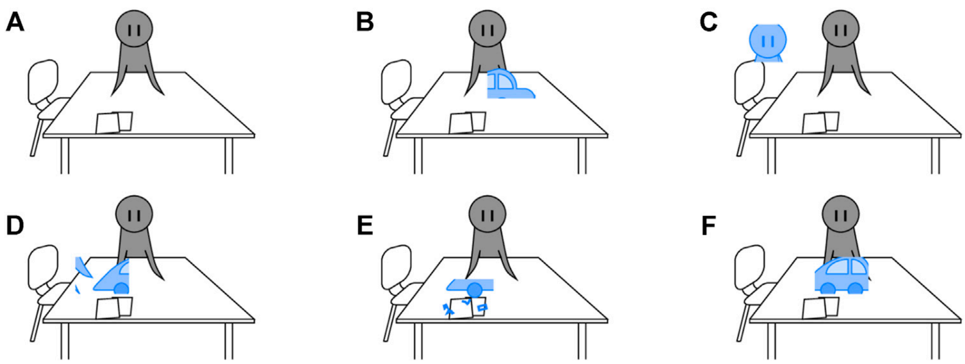

FIGURE 2 | Illustrations of a user's view when using an AR headset with limited FOV. Virtual content is shown in blue. In this figure, there is no rectangle indicating the virtual field of view in order to more accurately show the user's view of the scene. **(A)** The user is making eye contact with a physically co-located collaborator. **(B)** The user looks to part of the virtual car model that the collaborator is speaking about. The whole model does not fit into view, so only part is visible. **(C)** The user looks over at a remote collaborator. The virtual car is no longer visible. **(D)** The user looks down at the part of the car the remote collaborator is gesturing towards. The collaborator's face is no longer visible. **(E)** The user checks notes, showing part of the virtual car above the virtual annotates made on the paper notes. **(F)** The user tries to view the entire car, but it does not fit within the headset's FOV. To see the entire scene at once, see **Figure 1** panel B.

The horizontal range is unobstructed by the device, i.e., a real-world object at the far left or far right of the visual field is visible with or without the headset. Therefore, the horizontal field of view is approximately 180°, the horizontal field of view of the human visual system (Jones et al., 2013). In all, the unaugmented periphery of a user wearing the Microsoft HoloLens extends about 20° above the virtual FOV, 60° below it, and 75° to each side. While other headsets have a larger field of view, such as the Microsoft HoloLens 2 and the Magic Leap One, these headsets still have a sizable unaugmented periphery.

The unaugmented periphery raises some questions for AR researchers, designers, and users. How does the brain perceive virtual objects in the unaugmented periphery? In some way, this process is like perceiving an occluded object, which is an everyday occurrence. How does this process extend to cases in which the obscuring object is not a visible one, but rather is the invisible edge of some display? Furthermore, what does this odd kind of invisibility imply for presence of the virtual object within the unaugmented periphery? In this paper, we investigate this by intentionally placing a virtual human in this unaugmented periphery during a task, and we collect data on its social presence.

The task chosen for this work is the social facilitation and inhibition study in Miller et al. (2019) In the experiment, participants solved word puzzles at one of two levels of difficulty either being 'watched' by a virtual person or with no virtual person present. By including both of these conditions in our study, in addition to a new condition of the virtual person being inside the resting unaugmented periphery, we can not only test a new hypothesis, but also perform a replication of this work.

PREVIOUS WORK

Unaugmented Periphery

The term *unaugmented periphery* is introduced by Lee et al. (2018) to refer to the area a user can see real-world objects but cannot see virtual objects, i.e., the area within the real-world FOV but not within the virtual FOV. In the study, Lee and colleagues explored differences between a restricted FOV, produced by blocking out the unaugmented periphery with opaque foam, and an unrestricted FOV. Participants walked between two locations about 6 m apart. In the middle of the two locations was an obstacle, either a real or virtual person. They found that a participant's walking path around a real person was more similar to the walking path around a virtual person by a participant with restricted FOV than to the walking path around a virtual person by a participant with unrestricted FOV, which was interpreted as behavioral evidence that the restricted FOV causes the virtual person to have greater co-presence.

Perception of Invisible Objects

The sensory processing of vision information into coherent and persistent objects has been a subject of study in cognitive psychology. Object permanence, a concept rooted in Piaget (1954), is a person's understanding that objects can remain in existence even though the person does not receive any sensory stimulation from the object. The natural next step in this line of work involves determining the conditions and causes in judging an object as permanent rather than transient.

We bring two theories to bear upon the user's experience of the unaugmented periphery. In the first, we refer to work on the perceptual system and the importance of the visual patterns in the moments before the object ceases to be visible. In the second, we refer to a theory of presence that frames presence as the result of "successfully supported actions."

Exit Transition as Evidence of Existence

Gibson et al. (1969) suggest the distinction between an object out of sight or out of existence is made based on whether the object's final moments within view seem reversible. Reversible exits correspond to the object going out of sight, but non-reversible transitions correspond to objects going out of existence.

For example, picture an observer, Alice, in a hall watching Bob step into a room and close the door behind him. The moment Bob is no longer visible to Alice is the moment the door shuts. While Bob is closing the door, the sensory information that Alice receives indicating Bob's existence is the portion of him that is visible behind both the doorframe and the door. As Bob is closing the door, that portion becomes thinner and thinner, and the edge of the occluding objects (the door and the doorframe) stay consistent. If this scene is played backwards, it is just as realistic from Alice's point of view. Instead of the portion in which Bob is visible becoming smaller and smaller, it becomes larger and larger. This reversed scene would be visually very similar to Alice's view of Bob opening the door. This plausibility of the scene played backwards is what Gibson specifies as the key perceptual difference in the brain's conclusion of whether the object still exists.

As an example of a transition out of existence, consider a piece of newspaper burning up. If this scene is reversed in time, a newspaper would seem to appear from ashes. This implausible situation, according to Gibson's theory, is a signal to the perceptual system that the object has gone out of existence, not merely out of sight.

The application of this theory to the unaugmented periphery is a straightforward one. The pattern of visual information from the virtual human follows the pattern of occlusion, which is reversible. Therefore, objects in the unaugmented periphery would be perceived as going out of view rather than out of existence.

The fact that the occluding object is invisible does not invalidate this theory's application. Invisible occluding objects were discussed in Gibson's original work, driven by the foundational work of Michotte et al. (1964) and have been the subject of following studies, e.g. (Scholl and Pylyshyn, 1999).

Successfully Supported Actions as Evidence of Existence

A primary construct in the psychological study of virtual and augmented reality is *presence*, which we follow Lee (2004) in defining as the "perception of non-mediation." One theory of presence claims "successfully supported actions" (Zahorik and Jenison, 1998) are the root of presence. When actions are made towards an object, the object reacts, in some fashion, to the action made. When the object's response is congruent with the person's expectations of the response, the action is said to be successfully supported.

A very simple example of a successfully supported action is the counter-rotation of a virtual object when the user rotates their head. If a user rotates their head left to right, an object in front of them moves, relative to their field of view, from right to left. A second example would be a glass tipping over when bumped by a user's hand. The action is the hand contacting the glass, and the support is the production of a realistic tipping over effect.

Recent work in augmented reality aligns with this theory of presence being the result of "successfully supported actions." Work by Lee et al. (2016) placed participants at one end of a half-virtual, half-physical table. At the virtual end of the table was a virtual human interviewing the participant. The independent variable of the study was whether the wobbling of the table would be coherent between the virtual and physical worlds. In the experimental condition, if either the participant or the virtual character leaned on the table, the physical side of the table wobbled as if connected to the virtual side. This experimental condition resulted in higher presence and social presence of the virtual human than the control condition. In a study by Kim et al. (2018), co-presence of a virtual human was higher in the condition when virtual content (specifically, a person, some sheets of paper, and window curtains) in an AR scene to respond to the airflow from a real oscillating fan within the experiment room. In sum, realistic responses from virtual objects tend to increase presence.

When an object is within the unaugmented periphery, there is not only a lack of stimuli indicating existence, but also a violation of the user's expectations as to the virtual object's behavior, which are unsuccessfully supported actions. Because objects in front occlude objects in back, one expects that the virtual human should occlude the real-world objects behind it; this expectation is violated when the virtual human is virtually occluded in the unaugmented periphery. In sum, this theory of presence would predict objects in the unaugmented periphery to be less present than virtual objects within the virtual FOV.

Social Facilitation and Inhibition

Social facilitation and inhibition are complementary findings explaining the effect of an audience improving or impairing the performance of a task, depending on other contextual factors such as difficulty (Bond and Titus, 1983; Aiello and Douthitt, 2001). The earliest scientific mention of the social facilitation effect is from Triplett (1898) who found that children winding fishing reels of string in competition performed faster than those who wound alone.

These seemingly contradictory effects were synthesized into one theory by Zajonc (1965), who suggested that the effects were both due to the presence of others increasing arousal, and arousal increasing the likelihood of the dominant response. The bidirectional effects were due to the nature of the task. In the case the task was simple, and the dominant response was likely correct, an audience would improve performance. However, when the task was complex, the dominant response would be incorrect, and so an audience would impair performance.

The effects of social facilitation and inhibition can even occur due to the implied presence of others (Dashiell, 1930) or to virtual others (Hoyt, Blascovich and Swinth, 2003; Park and Catrambone, 2007; Zanbaka et al., 2007). In augmented reality, Miller et al. (2019) found evidence of a social facilitation and inhibition effect due to the presence of a virtual person displayed in augmented reality.

Current Work

The current work builds upon work by Lee et al. (2018) and Miller et al. (2019) s, and so we take some space here to provide a more direct contrast with these works.

Our study design is similar to the work of Miller et al. (2019). In the study of social facilitation an inhibition, the first of the three reported in that paper, participants performed a cognitive task in two conditions, with or without a virtual person. That study did not investigate the unaugmented periphery. In contrast, in the current study, we examine both conditions from the previous work in addition to a third condition in which the virtual human is intended to be within the unaugmented periphery.

While Lee et al. (2018) do focus on the unaugmented periphery and use behavioral measures of presence, their contribution is the proposal and test of a technological solution for the challenges raised by the unaugmented periphery. Specifically, they proposed to reduce the real-world FOV to increase presence. While effective as a piecemeal solution, this approach leaves many unanswered questions, most prominently the perceptual status of a virtual person within the unaugmented periphery in contrast to a fully present or fully absent virtual person.

In sum, this work investigates the perception of virtual humans within the unaugmented periphery to better ground discussions of its importance and potential solutions.

METHODS

Apparatus

The augmented reality display device was the Microsoft HoloLens headset, version 1. The field-of-view is 30° by 17.5°, and the resolution is 1,268 × 720 pixels for each eye. The device tracked its own position and orientation relative to the room and exported the data to a tracking file. The device also recorded the audio spoken by the participant during the task.

The cognitive task performed by the participant during the experiment was the anagram task used in both Park and Catrambone (2007) and Miller et al. (2019). Each anagram had five letters, and each participant was given ten anagrams to solve. Anagrams were broken up into easy and hard sets.

This study was performed in two locations to create a sufficiently large sample size. While this aggregation may reduce sensitivity, it also increases generalizability. The first was a booth at a science and technology museum in a large American city. This booth was approximately 2.5 m by 4 m. While seeing into the booth was not possible when the door was shut, the top of the booth was open and other museumgoers could easily be heard. In this location, the virtual person was 2.06 m from the participant on average.

The second location was a study room on a medium-size, suburban private college campus, 5.6 m by 6.4 m. The layout of the room is depicted in **Figure 3**. This room was separated from other rooms, and while there was a window connecting this room to another, the experimenter ensured the other room was empty. The virtual person was 3.83 m from the participant on average.

The virtual content displayed to the participants consisted of a Rocketbox Virtual Character (Gonzalez-Franco et al., 2020). To reduce gender effects, the virtual human displayed to the participant was the same gender as the participant. When the researcher pressed a button, the virtual human spoke a 20 s recorded introduction with talking gestures. When not speaking, the virtual human idled using a looped animation with a small amount of head motion and tilt.

Variables

The study design was preregistered at the Open Science Foundation[1], which included independent and dependent variables as well as covariates. Preregistering variables aims to reduce selective reporting of results based upon statistical significance (Nosek et al., 2018).

Independent Variables

The study design was a 3 × 2 design, with three conditions of visibility and two conditions of task difficulty. Each participant only experienced one of these combinations, making both variables between-subjects.

Visibility of Virtual Human

The virtual human can be Social, Outside FOV, or Alone. In the Social condition, the virtual human is present and placed near the anagram poster. The placement was about five degrees horizontally from the poster, and was chosen so that the virtual human would be within the field of view during the experiment. In the Outside FOV condition, the virtual human was present, but was farther away from the anagram poster, about 25°. This value was chosen so that the virtual human would be

[1]https://osf.io/h2ug3/.

FIGURE 3 | The left panel is a bird's-eye diagram of the study setup. The right panel is a photo of the campus-based study location, with the researcher posing in the place of a participant. The white dotted area indicates the user's virtual FOV.

outside the field of view, unless the participant looked in the direction of the virtual human. Note that the virtual human was visible as part of the pretest procedure even in the Outside FOV condition. Finally, there was the Alone condition, in which no virtual person was introduced or visible.

Anagram Difficulty

In order to test social facilitation and inhibition, rather than the virtual human changing performance across the board, there was two separate levels of difficulty, labeled as Easy and Hard. The anagram sets are the same ones as used in Miller et al. (2019). Each set consisted of 10 anagrams.

Dependent Variables

Social Presence

Social presence was measured as the average of the five-item Social Presence Questionnaire as used in Miller et al. (2019) with questions such as "To what extent did you feel like Chris was watching you?". These questions are included in the supplemental material. The scale gives five verbal options "Not at All" to "Extremely" that are mapped to integer values 1 to 5. Social presence values in this study had mean of 2.03, a standard deviation of 0.88, and a range from 1 to 4.6.

Anagram Score

The primary measure for measuring social facilitation and inhibition is the number of anagrams solved in the 3 minutes given for completing the task. The range of possible values was 0–10, as there were 10 anagrams to solve on each poster. On average, participants solved 5.92 anagrams, with a standard deviation of 3.19, and a range from 0 to 10.

Look-At Time

Based upon a definition of presence as successfully supported action, a virtual human that is visible more often would be predicted to have higher presence, as visibility can be thought of an action that is successfully supported.

To calculate whether the virtual human was visible or not, we calculated the central angle around the head between the headset's forward vector and the vector going from the headset to the virtual human. In short, this value measures the angular distance between the center of the field of view of the headset and the center of the virtual human. We chose 20° as the cut-off point such that the moments of time when the central angle was less than 20°, the virtual human was assumed to be visible, and the moments when the central angle was greater than 20°, the virtual human was assumed not to be visible. The final value used in statistics was the total time in seconds for which the virtual human was visible.

Procedure

Participants began the study with an online pre-screening, disallowing participants with epilepsy or high susceptibility to motion sickness. For participants on campus, this pre-screening was done two to 7 weeks before the study. For participants at the museum, this screening included in the consent form and explicitly mentioned to participants. At both locations, participants over 18 filled out the consent form, while participants younger than 18 completed an assent form while a parent or legal guardian completed the parental consent form. Participants then were asked to complete a pre-survey.

The training phase began with an explanation and test in solving the anagrams. Participants were prompted with two-word puzzles and were given about 40 seconds to complete them. If they did not, the experimenter informed them the answers and confirmed the answers made sense in the context of the instructions. Then, the experimenter introduced the HoloLens, including instructions for headset fit. At this point, we differ from the method in Miller et al. (2019) and do not have the participant walk towards virtual objects. This was due to the space constraints at the museum. At this point, if the participant was in either the Social or Outside FOV conditions, the virtual human was visible and introduced himself or herself to the participant. Participants in the Alone condition did not see the character, hear the introduction, or otherwise receive sensory information suggesting the presence of a virtual character.

The testing phase occurred when the experimenter placed a poster of anagrams visible to the participant and started the 3 min timer. At this point the experimenter stepped out of the room and waited to return. Upon return, the experimenter collected the headset and asked the participant to complete a post-experiment survey. The text of this survey is included in the supplementary

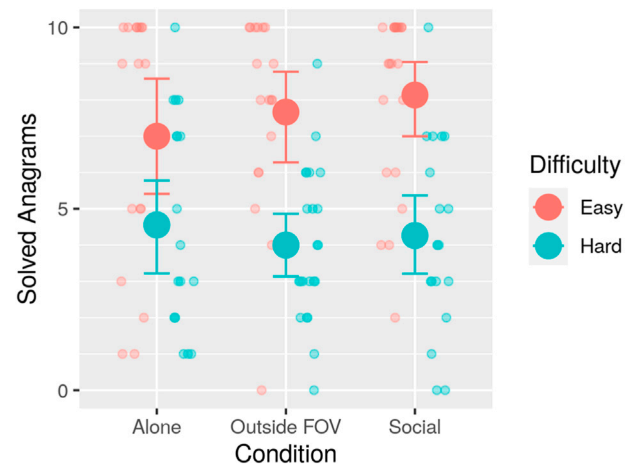

FIGURE 4 | Larger points represent means; smaller transparent points represent individual entries. Error bars are 95% bootstrapped confidence intervals of the mean. While the trends of social facilitation and inhibition are present, i.e., more easy anagrams and fewer hard anagrams are solved in the social condition than in the alone condition, the difference is not significant.

material. Finally, the participant was debriefed about the purpose and conditions of the study, and participants in the Alone condition were offered a chance to see the virtual person.

Participants

A total of 128 participants were collected under a university approved IRB protocol Twelve participants were not included in the analysis: three were not recorded, two recordings stopped halfway through, two elected to end early and leave, two participants tracking data were lost, two participants were not clearly informed of the anagram instructions, and one participant was not given the pre-survey. In total, the data of 116 participants (51 female, 64 male, and 1 nonbinary) could be analyzed. There were 29 participants from the museum, whose ages had a mean of 37.28, a standard deviation of 18.65, and a range from 13 to 81, and there were 87 participants on campus, whose ages had a mean of 20.34, a standard deviation of 1.42, and a range from 18 to 25. Breaking down by virtual human visibility and difficulty, there were 35 participants in the Alone condition (17 easy, 18 hard), 40 in the Outside FOV condition (18 easy, 22 hard), and 41 in the Social condition (22 easy, 19 hard).

RESULTS

In this study, we follow the preregistration distinction between confirmatory and exploratory tests (Nosek et al., 2018). Clarifying which kind of test is being run solidifies interpretations of confirmatory *p*-values while also allowing unexpected results to be reported. According to the preregistration, the primary confirmatory test is social facilitation and inhibition, which is an interaction effect between difficulty and visibility of the virtual human. This test is followed by others related to social presence and visibility.

Confirmatory Results

The confirmatory test in this study is an attempt of a replication of the first study in Miller et al. (2019) showing an effect on task performance based upon the interaction of task difficulty and virtual human visibility. This effect is interpreted as a social facilitation and inhibition effect. All three conditions (Alone, Outside FOV, Social) are given in **Figure 4**.

The statistical analysis performed was to predict the number of anagrams solved based on difficulty and visibility, with visibility limited to merely the Alone and Social conditions. The model was an ordinal logistic regression using the "polr" function from the "MASS" package for the R programming language. This was performed rather than a linear model due to the non-normality of the residuals indicated by the Shapiro-Wilk test ($W = 0.951$, $p = 0.005$).

The difference in anagram score due to difficulty was significant [$t\ (72) = -2.55$, $p = 0.013$, $r = -0.29$] such that participants solved fewer Hard anagrams than Easy anagrams. The visibility of the virtual human did not have a significant effect on anagram score [$t\ (72) = 0.863$, $p = 0.390$, $r = 0.10$]. The interaction effect between difficulty and visibility, which is our predicted indicator of social facilitation and inhibition, was not significant either [$t\ (72) = -0.698$, $p = 0.488$, $r = -0.08$]. For reference, the test was sensitive to an effect size of $r = +/-0.35$ at a power of 0.8.

Exploratory Results

Other tests and results explore the consequences of virtual humans outside the device's field of display. Because of their exploratory nature, these tests should encourage future confirmatory work.

Visibility Affects Social Presence

While the measure of anagrams solved was not sensitive enough to capture the social effect of others, the social presence questionnaire was able to. This questionnaire asked participants how socially present the virtual human felt to them during the experiment. Only participants in the Outside

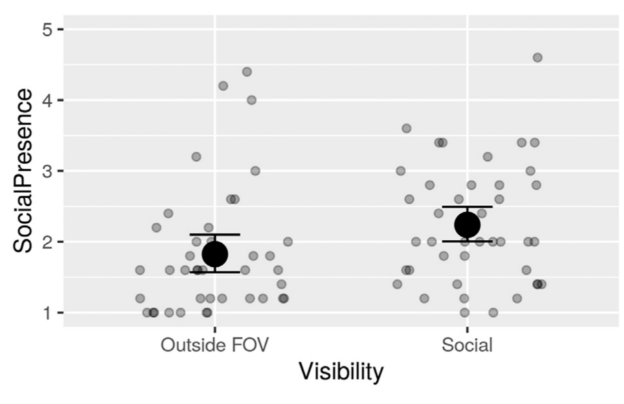

FIGURE 5 | Self-reported social presence was significantly higher in the Social condition than the Outside FOV condition. Larger points represent means; smaller transparent points represent individual entries. Error bars are 95% confidence intervals of the mean.

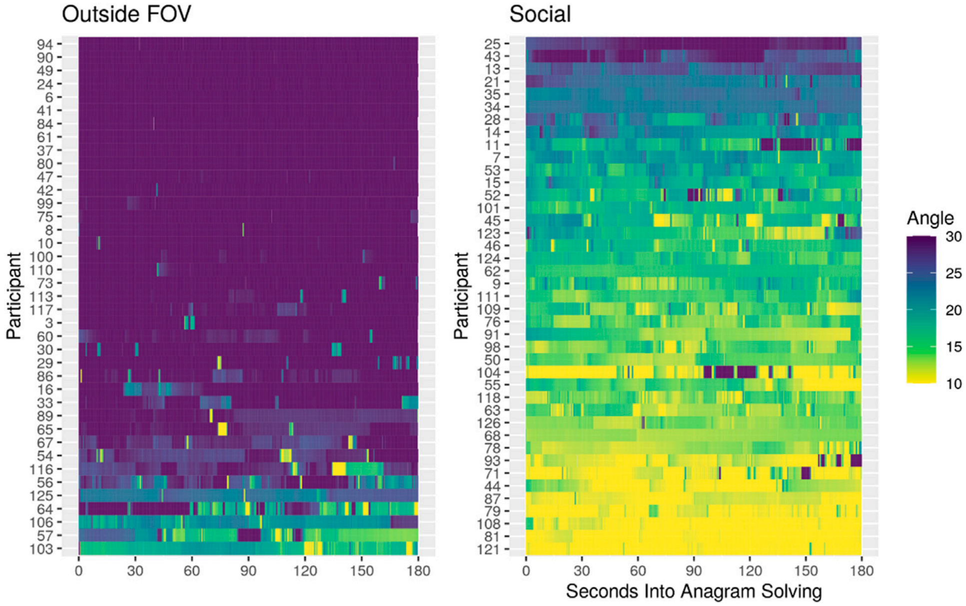

FIGURE 6 | Angle between head-forward and direction of virtual human for each participant in the study over time.

FOV and Social cases were asked this question, as there was no analogous question to ask when there was no virtual person. The results of this survey are plotted in **Figure 5**.

The statistical analysis performed was to predict the social presence rating based on visibility, with visibility limited to the Social and Outside FOV conditions. The test was a Kruskal-Wallis rank sum test from the base library for the R programming language, as the linear test had non-normally distributed residuals. The effect of visibility on social presence was significant (χ^2 (1, 79) = 6.87, p = 0.009, r = 0.27) such that participants in the Outside FOV condition reported less social presence than participants in the Social condition. This result validates intuition that when the virtual human is intentionally placed outside the device's field of view, the participants feel the virtual human is less socially present.

Visibility and Look-At Time

Gaze behavior of the participant is a useful behavioral variable. For the purposes of analysis, we have collapsed it into a single value, specifically, the amount of time the virtual human is in view while the participant is solving the anagrams. In this section, we wish to show the gaze behavior in finer detail. This serves two purposes: first, due to the few studies that investigate gaze behavior of objects within the unaugmented periphery, this can provide a global, intuitive sense of gaze behavior in this situation. Second, individual-level reporting of variables reveals unique features of participants (Molenaar and Campbell, 2009; Ram, Brose and Molenaar, 2013) and can provide opportunities to direct future lines of research.

Figure 6 displays how far, in degrees, the virtual human is from the center of the headset's field of view. In the plot, the angle values are clamped between 10 and 30° to focus on a more expressive range of values. This means that any values less than 10 are displayed as 10, and any values more than 30 are displayed as 30. These values were chosen because at a difference of 10°, the virtual human is certainly visible, and at 30°, the virtual human is certainly not visible.

Looking at the graph, the first visual feature is a strong visual difference in color between the two conditions. Because color represents visibility of the virtual human, this difference is interpreted as a manipulation check that participants did tend to focus on the anagram poster, and this made the virtual human in the Outside FOV condition usually not visible (>30°) while the virtual human in the Social condition was visible (<10°). A *t*-test between conditions confirms this (t (79) = -8.90, $p < 0.001$) Some participants never look back at the virtual human in the Outside FOV condition (e.g., 47, 49, 84), while others make a point to be looking at the virtual person (e.g., 57, 103). Most participants in this condition look at the virtual human only a few seconds (e.g., 29, 30, 86, 113).

DISCUSSION

The primary finding of this experiment is exploratory evidence that social presence of virtual humans in the unaugmented periphery is less than virtual humans in the augmented center. These findings are in accordance with a "successfully-supported-action" theory of presence.

The second finding is for augmented reality experiences. Application designers may expect some level of "curiosity" from users, e.g., to look around at other objects even when focusing on a task. This expectation should be calibrated against these results, which show that a portion of participants in the Outside FOV condition (16 of 40) never looked back at the virtual person once starting the task.

Finally, we speak of the lack of replication of the social facilitation and inhibition effect. With non-replications, there are points that almost always can be made, such as the true effect is smaller than the originally reported effect. We may also propose a few other reasons the effect did not replicate. Because many in this participant pool participated in previous AR studies, this population is not as naive as in previous work. According to our survey questions, 42 of 116 participants had previous experience with AR. Furthermore, one difference in the procedures between Miller et al. (2019) and this work is the pre-study interaction with AR content. The anagram-solving portion of the study does not require any physical movement on the participant's part, so when the avatar is present the experience may feel like a traditional screen-based character. Physical movement has been known to be a correlate of presence in some situations (Slater et al., 1998; Markowitz et al., 2018). In the original study, there was some minimal interaction with virtual content (walking towards virtual shapes) that could have increased presence due to successfully supported actions.

CONCLUSION

In this work, we present mixed results on the social presence of a virtual person outside the field of view. Participants who did not see the virtual person much, i.e., the virtual person most resided inside the unaugmented periphery, rated the virtual person as less socially present, but the behavioral measure of a word puzzle task was not significantly different. In addition, we attempt to replicate the first study in Miller et al. (2019) and do not find evidence of the same effect, but we do note the trends are in the same direction as originally reported.

Limitations

One limitation of this study is the sample size. While the subject size is twice as large as Miller et al. (2019), it still may be too small to find an accurate measure of effect size, considering the test was sensitive to effect sizes of r = 0.35 or larger. In addition, while there are some subjects from the general public, the majority of subjects are still college-age participants. Findings among users with different levels of comfort with digital devices may be different. In addition, augmented reality is not part of everyday experience in the locale this study was performed in, so there may be differences between effects occurring today and in the future due to novelty.

It is also worth noting that the two separate study locations (museum and campus) may have increased the variance or affected the looking behavior. While ideally this effect should be robust across locations and populations, it may have influenced the power of the results.

For the experimental design, it must be noted that the task did not involve any interaction with the virtual human and the virtual human had low behavioral realism (Blascovich, 2002). This difference must be taken into context when application designers consider interaction with characters in the unaugmented periphery.

Future Work

Questions remain towards the nature of the social facilitation and inhibition effect in augmented reality. If this effect is not present or significantly weaker than in virtual reality, then the features between the two media should be explored. For example, visual quality does not matter much for presence in virtual reality (Cummings and Bailenson, 2016), but with the juxtaposition of virtual content with the realism of the real world, there may be a much higher threshold in visual realism for augmented reality.

For the continuation of the investigation of unaugmented periphery, two variables we propose to investigate are varying modes of interaction and degrees of realism. For example, one may expect virtual characters that can interact verbally to be more socially present than those that interact only visually, especially when the character in the unaugmented periphery.

Studying the unaugmented periphery relates back to questions of perception and opens questions about presence in augmented reality. Better understanding of these unique situations of augmented reality will lead to better designed experiences in the future.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Stanford Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

JB and MM conceptualized the study. MM coordinated and performed the experiment. MM drafted the text and JB and MM edited substantially. JB provided study funding.

FUNDING

National Science Foundation, grants 1800922 and 1839974. The funders had no role in the study design, collection, analysis, or interpretation of the results.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frvir.2021.656473/full#supplementary-material

REFERENCES

Aiello, J. R., and Douthitt, E. A. (2001). Social Facilitation from Triplett to Electronic Performance Monitoring. *Group Dyn. Theor. Res. Pract.* 5, 163–180. doi:10.1037/1089-2699.5.3.163

Azuma, R. T. (1997). A Survey of Augmented Reality. *Presence: Teleoperators & Virtual Environments* 6, 355–385. doi:10.1162/pres.1997.6.4.355

Blascovich, J. (2002). *Social Influence within Immersive Virtual Environments*, 127–145. doi:10.1007/978-1-4471-0277-9_8

Bond, C. F., and Titus, L. J. (1983). Social Facilitation: A Meta-Analysis of 241 Studies. *Psychol. Bull.* 94, 265–292. doi:10.1037/0033-2909.94.2.265

Cummings, J. J., and Bailenson, J. N. (2016). How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence. *Media Psychol.* 19, 272–309. doi:10.1080/15213269.2015.1015740

Dashiell, J. F. (1930). An Experimental Analysis of Some Group Effects. *J. Abnormal Soc. Psychol.* 25, 190–199. doi:10.1037/h0075144

Gibson, Js. J., Kaplan, G. A., Reynolds, H. N., and Kirk, W. (2019). The Change from Visible to Invisible: A Study of Optical Transitions*. *Perception & Psychophysics* 5, 194–202. doi:10.4324/9780367823771-19

Gonzalez-Franco, M., Ofek, E., Pan, Y., Antley, A., Steed, A., Spanlang, B., et al. (2020). The Rocketbox Library and the Utility of Freely Available Rigged Avatars. *Front. Virtual Real.* 1, 1–23. doi:10.3389/frvir.2020.561558

Hoyt, C. L., Blascovich, J., and Swinth, K. R. (2003). Social Inhibition in Immersive Virtual Environments. *Presence: Teleoperators & Virtual Environments* 12, 183–195. doi:10.1162/105474603321640932

Jones, J. A., Swan, J. E., and Bolas, M. (2013). Peripheral Stimulation and its Effect on Perceived Spatial Scale in Virtual Environments. *IEEE Trans. Vis. Comput. Graph* 19, 701–710. doi:10.1109/TVCG.2013.37

Kim, K., Bruder, G., and Welch, G. F. (2018). "Blowing in the Wind: Increasing Copresence with a Virtual Human via Airflow Influence in Augmented Reality," in *International Conference on Artificial Reality and Telexistence Eurographics Symposium on Virtual Environments*. Editors G. Bruder, S. Cobb, and S. Yoshimoto.

Kreylos, O. (2015). *On the Road for VR: Microsoft HoloLens at Build 2015*, 5. San Francisco. Available at: http://doc-ok.org/?p=1223 (Accessed May 3, 2021).

Lee, K. M. (2004). Presence, Explicated. *Commun. Theor.* 14, 27–50. doi:10.1111/j.1468-2885.2004.tb00302.x

Lee, M., Bruder, G., Hollerer, T., and Welch, G. (2018). Effects of Unaugmented Periphery and Vibrotactile Feedback on Proxemics with Virtual Humans in AR. *IEEE Trans. Vis. Comput. Graphics* 24, 1525–1534. doi:10.1109/TVCG.2018.2794074

Lee, M., Kim, K., Daher, S., Raij, A., Schubert, R., Bailenson, J., and Welch, G. (2016). "The Wobbly Table: Increased Social Presence via Subtle Incidental Movement of a Real-Virtual Table," in Proceedings - IEEE Virtual Reality, 2016-July: 11–17, Greenville, South Carolina. doi:10.1109/VR.2016.7504683

Markowitz, D. M., Laha, R., Perone, B. P., Pea, R. D., and Bailenson, J. N. 2018. "Immersive Virtual Reality Field Trips Facilitate Learning about Climate Change." *Front. Psychol.* 9, 1–20. doi:10.3389/fpsyg.2018.02364

Michotte, A., Thines, G., and Crabbe, G. (1964). Amodal Completion of Perceptual Structures. *Michotte's Experimental Phenomenology of Perception*. New York, NY: Routledge.

Miller, M. R., Jun, H., Herrera, F., Yu Villa, J., Welch, G., and Bailenson, J. N. (2019). Social Interaction in Augmented Reality. *PLOS ONE* 14, e0216290. doi:10.1371/journal.pone.0216290

Molenaar, P. C. M., and Campbell., C. G. (2009). The New Person-specific Paradigm in Psychology. *Curr. Dir. Psychol. Sci.* 18, 112–117. doi:10.1111/j.1467-8721.2009.01619.x

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., and Mellor, D. T. (2018). The Preregistration Revolution. *Proc. Natl. Acad. Sci. USA* 115, 2600–2606. doi:10.1073/pnas.1708274114

Park, S., and Catrambone, R. (2007). Social Facilitation Effects of Virtual Humans. *Hum. Factors* 49, 1054–1060. doi:10.1518/001872007X249910

Piaget, J. (1954). *The Construction of Reality in the Child*. Basic Books.

Ram, N., Brose, A., Peter, C., and Molenaar, M. (2013). "Dynamic Factor Analysis: Modeling Person-specific Process," in *The Oxford Handbook of Quantitative Methods (Vol 2): Statistical Analysis*. New York, NY: Oxford University Press, 441–457. Avaialable at: http://www.redi-bw.de/db/ebsco.php/search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2013-01010-021&site=ehost-live. doi:10.1093/oxfordhb/9780199934898.013.0021

Scholl, B. J., and Pylyshyn, Z. W. (1999). Tracking Multiple Items through Occlusion: Clues to Visual Objecthood. *Cogn. Psychol.* 38, 259–290. doi:10.1006/cogp.1998.0698

Slater, M., McCarthy, J., Maringelli, F., and Maringelli, F. (1998). The Influence of Body Movement on Subjective Presence in Virtual Environments. *Hum. Factors* 40, 469–477. doi:10.1518/001872098779591368

Triplett, N. (1898). The Dynamogenic Factors in Pacemaking and Competition. *Am. J. Psychol.* 9, 507–533. doi:10.2307/1412188

Zahorik, P., and Jenison, R. L. (1998). Presence as Being-In-The-World. *Presence* 7, 78–89. doi:10.1162/105474698565541

Zajonc, R. B. (1965). Social Facilitation. *Science* 148, 269–274. doi:10.1037/e505132009-001

Zanbaka, C. A., Ulinski, A. C., Goolkasian, P., and Hodges, L. F. (2007). "Social Responses to Virtual Humans," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 1561–1570. doi:10.1145/1240624.1240861

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

ORIGINAL RESEARCH
published: 27 September 2021
doi: 10.3389/frvir.2021.694453

Despite Appearances: Comparing Emotion Recognition in Abstract and Humanoid Avatars Using Nonverbal Behavior in Social Virtual Reality

*Yilu Sun and Andrea Stevenson Won**

Department of Communication, Cornell University, Ithaca, NY, United States

The ability to perceive emotional states is a critical part of social interactions, shaping how people understand and respond to each other. In face-to-face communication, people perceive others' emotions through observing their appearance and behavior. In virtual reality, how appearance and behavior are rendered must be designed. In this study, we asked whether people conversing in immersive virtual reality (VR) would perceive emotion more accurately depending on whether they and their partner were represented by realistic or abstract avatars. In both cases, participants got similar information about the tracked movement of their partners' heads and hands, though how this information was expressed varied. We collected participants' self-reported emotional state ratings of themselves and their ratings of their conversational partners' emotional states after a conversation in VR. Participants' ratings of their partners' emotional states correlated to their partners' self-reported ratings regardless of which of the avatar conditions they experienced. We then explored how these states were reflected in their nonverbal behavior, using a dyadic measure of nonverbal behavior (proximity between conversational partners) and an individual measure (expansiveness of gesture). We discuss how this relates to measures of social presence and social closeness.

Keywords: virtual reality, emotion perception, nonverbal communication, personality traits, emotional states, expansiveness of gesture, avatars, proximity

Edited by:
Missie Smith,
Independent researcher, Detroit, MI, United States

Reviewed by:
Richard Skarbez,
La Trobe University, Australia
Benjamin J. Li,
Nanyang Technological University, Singapore

****Correspondence:***
Andrea Stevenson Won
a.s.won@cornell.edu

Specialty section:
This article was submitted to Virtual Reality and Human Behaviour, a section of the journal Frontiers in Virtual Reality

Received: *13 April 2021*
Accepted: *10 August 2021*
Published: *27 September 2021*

Citation:
Sun Y and Won AS (2021) Despite Appearances: Comparing Emotion Recognition in Abstract and Humanoid Avatars Using Nonverbal Behavior in Social Virtual Reality.
Front. Virtual Real. 2:694453.
doi: 10.3389/frvir.2021.694453

1 INTRODUCTION

Perceiving emotional states is a critical part of social interactions, shaping how people understand and respond to each other. In face-to-face communication, people's perceptions of emotions depend on multiple factors including facial expressions (Montagne et al., 2007; Schirmer and Adolphs, 2017), and verbal (De Gelder and Vroomen, 2000, Doyle and Lindquist, 2017), and nonverbal behavior (Hertenstein et al., 2009; Enea and Iancu, 2016). As immersive technology becomes more prevalent, more interpersonal communication is occurring in social virtual reality platforms (McVeigh-Schultz et al., 2018). One important measure of user experience in social VR is thus the ability to perceive emotional states. In virtual reality, how an avatar appears and what behavior is tracked and rendered must be designed. This allows nonverbal behavior as well as other affective cues to be conveyed in ways that diverge from the usual human appearance. Thus, understanding the relationship between avatar appearance, how people recognize others' emotions and how they express themselves can help guide the design of expressive avatars and the user experience elicited.

Previous research has found that both the appearance and the behavior of avatars impacts participants' emotional reactions (Mousas et al., 2018). One study result showed that participants expressed different anxiety levels when giving a speech to virtual audiences depending on whether they presented negative, positive or neutral emotions (Pertaub et al., 2002). In another study, the rapport towards agent-avatars depended on an interaction between appearance and personality, meaning that under some circumstances, realism could be a better choice for avatars (Zibrek et al., 2018). The interaction of avatar realism and avatar appearance influences users' social perceptions (Garau et al., 2003; Thomas et al., 2017; Nowak and Fox, 2018), including in the social domain (Ehrsson et al., 2005; Bailenson et al., 2006; Lugrin et al., 2015; Latoschik et al., 2017). This paper makes several contributions to understanding emotion perception in avatars. First, we describe two types of avatars, one abstract and one humanoid, both of which convey important qualities of nonverbal behavior. Second, we examine emotion recognition during social interactions in virtual reality, comparing partners' accuracy with participants' self-reports. Third, we examine how tracked nonverbal behaviors relates to emotion recognition. Finally, we relate emotion recognition to other measures of user experience, including social presence and social closeness.

Below we review the literature on avatar appearance and behavior, describe our experimental setup and results, and discuss the implications of these findings for avatar design.

1.1 Avatar Appearance and Behavior

To interact with other people in immersive VR platforms, people need to be embodied with virtual avatars. This opens up opportunities for people to create or choose virtual representations that may or may not resemble themselves. Such avatars can have different levels of consistency with the user's own body, from realistic humanoid avatars that closely resemble the user's physical body to abstract shapes with no customized features (Bailenson et al., 2006). People can even break traditional norms of embodiment, such as magnifying nonverbal cues (Yee et al., 2008) or remapping the avatars' movement in a novel way (Won et al., 2015).

Avatar appearances not only reflect people's virtual identities and their self-perceptions, but also can impact how they behave and interact with others (Garau, 2003; Roth et al., 2016). For example, the *Proteus* effect demonstrated that embodiment in an avatar could lead the user to behave according to their expectations of that avatar (Yee and Bailenson, 2007). Different avatar appearances also have been found to yield different interaction outcomes (Latoschik et al., 2017). People have responded to anthropomorphic avatars more positively and have been more willing to choose an avatar that reflects their gender (Nowak and Rauh, 2005). Some research suggests that customized avatars significantly improve users' sense of presence and body ownership in virtual reality (Waltemate et al., 2018). However, due to the limitations of avatar creation and control, some avatar appearances may fall into the "uncanny valley," such that the representations, though closely resembling humans, are not quite real enough to be acceptable (Mori, 1970). For example, researchers found that avatar realism created "eeriness" that influenced people's accuracy to judge extroversion and agreeableness as personality traits (Shin et al., 2019). On the other hand, other research has found that avatars with human appearance elicit a slightly lower 'illusion of virtual body ownership' compared to machine-like and cartoon-like avatars (Lugrin et al., 2015). To continue these investigations we aimed to explore how avatars of varying realism that gave similar nonverbal information might affect emotion recognition.

1.2 Emotion Perception From Movement

Though facial expressions and voice are important channels for emotion recognition (Banse and Scherer, 1996), bodily movements and posture also convey critical emotional cues (Dael et al., 2012). Gesture is an integral part of nonverbal communication that conveys emotion (Dael et al., 2013). Furthermore, body movement could also be used to predict people's emotional ratings of others (De Meijer, 1989). By combining gestures and facial expressions, people were able to be more accurate in emotion recognition (Gunes and Piccardi, 2007), as people infer each other's emotions through channels that process implicit messages (Cowie et al., 2001). Emotions can even be identified from minimal information about body movements, as in the phenomenon "point light display" (Atkinson et al., 2004; Clarke et al., 2005; Lorey et al., 2012). First introduced by Johansson (Johansson, 1973), the phenomenon has been confirmed and evolved with new findings (Kozlowski and Cutting, 1977; Missana et al., 2015). For example, when body movements were exaggerated, recognition became more accurate and led to higher emotional ratings (Atkinson et al., 2004).

Researchers have found that Big Five personality traits are related to verbal and nonverbal behavior in VR (Yee et al., 2011). Biological motion also correlated to personality traits. For example, with minimal movement information that people relied on to make the first impression, people were able to predict others' perceived personality (Koppensteiner, 2013). Even when the personality traits were inferred from a thin slice of movement, there is a significant correlation with the information provided by knowledgeable informants (Borkenau et al., 2004). Understanding others' personality traits is important because it influenced how people felt about their interactions with others, and how they evaluated themselves and others in virtual reality (Astrid et al., 2010). Some personality traits that exhibited specific movement patterns were predictive of people's evaluations (Astrid et al., 2010). In terms of social interactions with a virtual character, a study showed that certain personality traits were found to be more easily inferred from the avatars and those avatars also revealed accurate information about specific personality traits of people who created them (Fong and Mar 2015). How much people like the virtual characters depended on an interaction between the virtual avatars' appearances and personalities (Zibrek et al., 2018). Thus, in our study, we asked participants to report on both emotional state and personality trait for themselves and their partners.

Researchers have proposed various ways to analyze behavior in virtual reality using the movement data provided by trackers in virtual reality headsets and hand controllers (Kobylinski et al., 2019). Using movement data, researchers found that people who performed a communication task better in virtual reality had more movement using their gestures as an aid to communicate

(Dodds et al., 2010). Additionally, when people interacted with humans or virtual agents with open gestures in a mixed reality platform, they were more willing to interact and they reported being more engaged in the experience (Li et al., 2018). Mimicry behavior, known as the "chameleon effect," showed that when people interacted with a virtual agent that mimicked them with a time lag, these agents received more positive ratings on their traits (Bailenson and Yee, 2005; Tarr et al., 2018). We thus ask how emotion can be perceived through participants' movements when they are rendered through an avatar, and whether a humanoid or abstract avatar appearance makes a difference if the movement data conveyed is equivalent.

1.3 Current Study

This paper describes the second part of a pre-registered study on nonverbal behavior and collaboration in immersive virtual reality (Sun et al., 2019). In the first paper from this study, we followed our pre-registered hypotheses and research questions to examine whether nonverbal synchrony would emerge naturally during conversations in virtual reality; if it would differ depending on the types of avatars participants used, and if nonverbal synchrony would be linked to task success. Stronger positive and negative correlations between real pairs compared to an artificial "pseudosynchrony" pairs were found supporting the hypothesis that nonverbal synchrony occurred naturally in dyadic conversations in virtual reality. Though there was no significant correlation between the task success and nonverbal synchrony, we found a positive significant correlation between social closeness and nonverbal synchrony. In this second paper, we address the remaining research question and add exploratory analyses to address the relationship between emotion recognition, individual and dyadic measures of nonverbal behavior (specifically, proximity and openness of gesture) and avatar appearance. We defined open gestures as the expansiveness of participants' hand movements, and operationalized this measure as the distance between the participants' hands. We asked participants to self-report their own emotional states, as well as to estimate their partners' emotional states. These state measures were not analyzed elsewhere, and we analyze them now for the first time to answer the final research question of that pre-registration:

RQ4: Will there be an effect of appearance on emotion perception, such that a conversational partner perceives emotion differently depending on whether participants are represented by a cube or a realistic-looking avatar?

In our study, we designed two conditions with different avatar appearances to convey approximately the same information about users' posture and gestures. One avatar was humanoid, and the other cube-shaped. Because we used consumer headsets, we were limited to tracking data derived from the position of the headset, and the position of the two hand controllers. In both the humanoid and cube avatars, the position of the avatar was linked to the head tracker. Similarly, we aimed to provide equivalent information about the position of the hands. The humanoid avatar hands followed the position of the participants' hand controllers. In the cube avatar, the sides of the cube scaled depending on where the participant held the hand tracker. Thus, we explored two measures of nonverbal behavior that were clearly related to this positional or postural information. One behavior was measured on the pair level: proximity, or the distance between the two participants in a pair (Won et al., 2018). The other measure was on the individual level: expansiveness of gesture, or how far apart an individual participant held his or her hands (Li et al., 2018).

In addition to answering this question, we conducted exploratory analyses on participants' tracked movements and related these to participants' self-reported emotional states.

Finally, we also collected participants' self-reported personality traits and their ratings of their partners as part of a larger on-going study. While we did not pre-register research questions for these data, we did duplicate the analyses for emotional states with these data. We present those results in the supplementary materials in **Appendix B**.

The complete data set can be found here (https://doi.org/10.6077/xvcp-p578) and is available on submission of an approved IRB protocol to the archive.

2 METHODS AND MATERIALS

2.1 Participants

This study was approved by the Institutional Review Board and all participants gave informed consent. We excluded one pair of participants due to motion sickness and one pair who reported being close friends or relatives. After removing participants with missing movement data, due to sensor issues, there were 76 pairs of participants left for the analyses.

Participants ($n = 152$) were 49 males, 102 females and one participant who preferred not to reveal their gender. Participants were randomly assigned to pairs, resulting in 33 male-female pairs, eight male-male pairs, and 34 female-female pairs, and one pair whose gender composition was not revealed. When describing their race/ethnicity, 12 participants described themselves as African Americans, 68 as Asian or Pacific Islanders, 69 as Caucasians as multi-racial, 7 people selected more than one race/ethnicity, eight people described themselves as "other" and three people chose "I prefer not to answer this question." Participants received course credits or cash compensation for the experiment. 41 pairs of participants were assigned to the humanoid avatar condition, and 35 pairs of participants to the cube avatar condition.

2.2 Apparatus

Participants wore Oculus Rifts and held Oculus Touch hand controllers. Movement data from these components were stored on a database on a local server. The experimental environment was created using the game engine Unity 3D.

2.3 Data Collection Procedure

A tracking platform saved the movement data from both participants at 30 times per second. Data timestamping was done on the movement tracking platform to avoid discrepancies arising from using each client's individual system

FIGURE 1 | The figure shows the third person view of two customized humanoid avatars in the brainstorming task.

clock. The setup for the tracking platform, its architecture, and the procedures used to minimize noise due to latency are described in detail in (Shaikh et al., 2018).

2.4 Experiment Flow

Each participant visited the lab twice. First, they visited the lab to be photographed. The photographs were taken in a passport style with no glasses and in neutral lighting.

Participants were then randomly assigned to one of two avatar appearance conditions; humanoid avatar and cube avatar. If participants were assigned to the humanoid avatar conditions, research assistants then created their avatars based on their photos, using the procedure described in Shaikh et al. (2018). In the humanoid avatar condition, participants could see their hands and bodies from the first person perspective. In the abstract cube condition, participants were represented by generic white cubes with no personalized features.

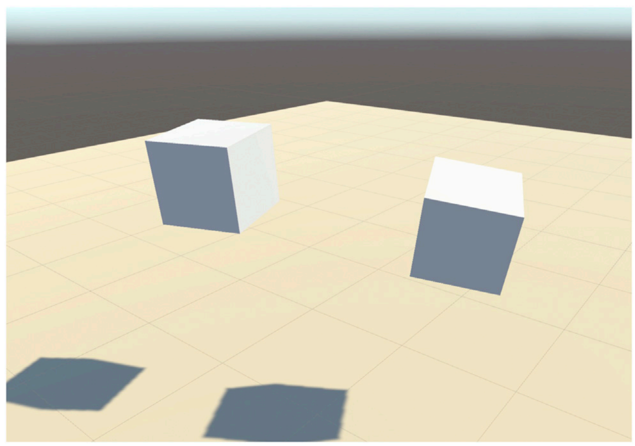

FIGURE 2 | The figure shows the third person view of two abstract cube avatars in the brainstorming task.

2.4.1 Second Visit: The Experiment

At their second visit, participants were instructed to go to two different lab rooms on different sides of the building so that they did not see each other in person before the experiment. A research assistant in each room assisted participants with the head-mounted display and the hand controllers. Participants completed the remainder of the task in a minimal networked virtual environment consisting of a plain white platform with directional lighting.

Participants first experienced their avatar body in a solitary mirror scene. They performed three exercises: raising up their arms, holding their arms wide and then folding them, and stepping toward and back from the mirror. These exercises helped the participants to gain a sense of embodiment in the assigned avatars.

In the humanoid avatar condition, the avatar's head and hands followed the movement of the head and hand trackers, while the rest of the avatar body was animated by inverse kinematics (Tolani et al., 2000). Participants could see avatars customized with their own faces in the first, mirror scene, but during the interaction, which occurred in a separate scene without a mirror, could only see avatar hands and the rest of their avatar bodies from the first person perspective.

In the abstract cube condition, the volume of the cube avatar shrank or grew as participants moved their hands. In other words, each cube got bigger or smaller as participants moved the hand controllers closer together or farther apart. In addition, the angle and position of the cube followed the angle and position of the participant's head. Participants could see their own shadows as well as the avatars and shadows of conversational partners.

After the mirror scene, participants were then connected to their conversational partners in a mirror-less environment. **Figures 1**, **2** show the humanoid avatar pairs and the abstract cube pairs. Participants then completed a brainstorming task, either competing or collaborating with their conversational partners. While participants' ideas were scored to answer earlier hypotheses, they are not further discussed in this paper. Collaborative and competitive conditions differed only in small variations in the verbal instructions to participants. For the purposes of this analysis, we collapse across conditions.

After the 5-min brainstorming task in VR, participants were directed to another laptop to complete a post-test survey.

2.5 Measures

Below, we list the measures of emotional accuracy. In response to this issue's call to reanalyze previous research from a new perspective, we also have reanalyzed the measures of presence found in Sun et al. (2019) as well as the Witmer and Singer immersion measures which were not previously analyzed.

2.5.1 Social Closeness

In order to have a proxy for rapport, we used a measure of social closeness that was previously collected but had not previously been used in emotion recognition analysis. Following previous work on social closeness, we asked 10 questions on liking, affiliation, and connectedness (Won et al., 2018) (*alpha* = 0.92). We averaged 10 questions for each individual, and averaged these with their partners' scores to create a pair-level social closeness measure (M = 3.429, SD = 0.586). Unlike the other measures discussed, this measure was previously correlated with nonverbal synchrony in the previous paper.

2.5.2 Emotional State

In the post-experiment questionnaire, participants were asked to rate to what extent the 18 emotional adjectives represent their current feelings and states. Then they were asked to rate how they guessed the same adjectives would represent their conversational partners' feelings and states. The rating was on a 4-point scale (1 = disagree strongly to 4 = agree strongly).

This questionnaire was adapted from the UWIST mood checklist (Matthews et al., 1990). The 18 adjectives were grouped into three categories: hedonic tone items, tense arousal items, and energetic arousal items. For each item there were three positive and three negative adjectives respectively. To process the data, we first reverse coded the ratings for the negative adjectives. Then we categorized the adjectives into three groups for each participant and took the average of the adjective ratings in that group to create three new variables: hedonic tone, tense arousal, and energetic arousal. For each individual, we thus had their self-rating of their emotional state, and their other-rating of their partner's emotional state. For each group, we created the group self ratings on hedonic (M = 2.439, SD = 0.587), tense (M = 1.823, SD = 0.548) and energetic (M = 2.965, SD = 0.595) states as well as the group partner ratings hedonic (M = 2.253, SD = 0.553), tense (M = 1.837, SD = 0.462) and energetic (M = 3.034, SD = 0.536) states by averaging both participants' self ratings and their ratings on their partners respectively.

To create the emotional consensus measure, we calculated the correlation between two participants in a pair's emotional ratings. First we calculated the correlation between Participant A's eighteen self emotional ratings and Participant B's emotional ratings to A. Then vice versa to get Participant B's eighteen self-ratings. The emotional consensus score for the pair is the average of these two ratings (M = 0.298, SD = 0.251).

There is at least one other way to score participants' accuracy on ratings of emotional state. This could be done by calculating the difference between the two participants' ratings, as follows. First we calculated the difference between Participant A's self emotional ratings and Participant B's emotional ratings of Participant A. Then vice versa to get Participant B's self-ratings and Participant A's ratings of Participant B. The pairwise emotional recognition score is thus the average of these two ratings, or how close each person's rating of their partner was to the "ground truth" of their partner's self-rating (M = 0.827, SD = 0.230). This method produces very similar results to the measure of emotional accuracy described above, so for brevity, we describe those outcome measures in **Appendix A**.

2.5.3 Personality Traits

In the post experiment questionnaire, the participants were also asked to choose the extent to which they agree or disagree with ten personality traits that might or might not apply to them. Participants were also asked to rate the same ten personality traits that may or may not apply to their conversational partners. As with the emotional state measures, these ratings were then calculated to create both individual and pair measures. The ratings were on a 7-point scale (1 = disagree strongly to 7 = agree strongly). The questionnaire was adapted from Ten Item Personality Measure (TIPI) (Gosling et al., 2003). For each individual, we thus had their self-rating of their personality traits, and their other-rating of their partner's personality traits.

To create the trait consensus measure, we calculated the correlation between two participants in a pair's personality traits ratings. First we calculated the correlation between Participant A's self-rating of personality traits and Participant B's rating of A's personality traits. Then, we repeated that calculation in reverse to get Participant B's ratings. The personality traits score for the pair is the average of these two ratings (M = 0.394, SD = 0.265).

2.5.4 User Experience/Presence

In the post-experiment questionnaire, the participants were asked 10 questions about their user experience and sense of presence in the virtual environment (Witmer and Singer, 1998). Based on the categorizations of those questions described in Witmer et al.

TABLE 1 | User experience and presence questionnaire.

Categories	Survey questions
Adaption/immersion	• How quickly did you adjust to the virtual environment experience? • How proficient in moving and interacting with the virtual environment did you feel at the end of the experience? • How much delay did you experience between your actions and expected outcomes?
Involvement	• How much did your experiences in the virtual environment seem consistent with your real world experiences? • How compelling was your sense of moving around inside the virtual environment? • How natural did your intentions with the environment seem? • How well could you concentrate on the assigned tasks or required activities rather than on the mechanisms used to perform those tasks or activities?
Visual Quality	• How much did the visual aspects of the environment involve you?
Audio Quality	• How much did the auditory aspects of the environment involve you?
Distraction	• How much did the visual display quality interfere or distract you from performing assigned tasks or required activities?
Self-presence	• If something happened to the avatar, it was happening to me • The avatar was an extension of me • The avatar represented my actions well • The avatar was me
Social Presence	• I felt like the other participant was present • I felt like I was in the same room with the other participant • I felt like the other participant was aware of my presence • I felt like the other participant was real

(2005), we broke these measures into five categories: adaptation/immersion, involvement, visual quality, audio quality and distraction. The detailed survey questions are listed in Table 1.

Adaptation/immersion

Due to the low internal consistency score for the three questions above (*alpha* = 0.51). The question "How much delay did you experience between your actions and expected outcomes?" was dropped for analysis, which increased the *alpha* to 0.69. For every participant in the pair, the adaptation/immersion score is the average of the two questions (M = 3.645, SD = 0.608).

Involvement

The questions about how "natural or compelling" the experience was are categorized based on Witmer et al. (2005). The internal consistency was low (*alpha* = 0.66). After dropping the question "How well could you concentrate on the assigned tasks or required activities rather than on the mechanisms used to perform those tasks or activities?", the *alpha* was increased to 0.72, and this subset was then used for the rest of the analysis. For every participant in the pair, the involvement score was the average of the three questions (M = 2.414, SD = 0.859).

Visual quality

For every participant in the pair, the visual quality score is the average of this question (M = 2.151, SD = 1.031).

Audio quality

For every participant in the pair, the audio quality score is the average of this question (M = 3.217, SD = 1.150).

Distraction

For every participant in the pair, the distraction score is the average of this question (M = 2.795, SD = 1.133).

We compared these measures to two other measures of presence that were discussed in Sun et al. (2019), self-presence (*alpha* = 0.84) and social presence (*alpha* = 0.82).

Then, the self-presence scores and the social presence scores from two participants in a pair were averaged respectively to get the group's self-presence scores (M = 2.277, SD = 0.656) and the group's social presence scores (M = 3.086, SD = 0.594).

3 RESULTS

Below, we report all analyses conducted on these measures. Some presence measures were examined on both the pair level (social closeness and social presence) while other measures were examined only on the individual level (self-presence and the Witmer and Singer immersion questions).

3.1 Emotion Recognition Accuracy and Avatar Appearance

We first sought to answer our final research question in our previous pre-registration link here: *RQ4: Will there be an effect of appearance on emotion perception, such that a conversational partner perceives emotion differently depending on whether participants are represented by a cube or a realistic-looking avatar?*

First, we explored whether there is a difference in how emotional states were perceived by participants depending on the appearance of the avatar used during their interaction. In other words, did participants who saw their partners represented by cubes rate their partner's emotional state as more tense, more energetic, or more hedonic than did participants who saw their partners represented by humanoid avatars?

We used a linear mixed-effect model in R's lme4 package, including the pair ID as a random effect to account for the non-independence of the two partner's ratings. We tested whether the appearance of the partners' avatar (both members of a pair always had the same avatar condition) predicted participants' ratings of their conversational partners' emotional states. We found no difference between participants' ratings of their conversational partners' hedonic ($F = 0.129$, $p = 0.721$), tense ($F = 0.009$, $p = 0.926$) or energetic ($F = 0.207$, $p = 0.650$) emotional states regardless of the avatar appearances of their conversational partners. In other words, avatar appearance did not impact participants' perceptions of their partners' emotional states.

Secondly, we explored whether participants who saw their partner in a humanoid avatar could perceive their partner's emotional states more *accurately* than those who saw their partner represented by an animated cube. To do this, we used the emotional consensus scores.

The emotional consensus score, which was generated by averaging the correlation between participants' ratings of their partners and their partner's rating of themselves, indicated that participants were able to recognize each other's emotional states at a rate significantly different than chance across both conditions ($M = 0.298$, $SD = 0.251$) using a one sample t-test comparing the result with 0 (t (71) = 10.077, $p < 0.001$).

Since the emotional consensus score is normally distributed ($W = 0.984$, $p = 0.520$) using a Shapiro-Wilk normality test, we used a Welch Two Sample t-test. And we found that there was no significant difference in the emotion consensus scores (t (66.093) = 0.155, $p = 0.878$) whether participants were represented by humanoid avatars or cubes. In other words, we did not see a difference between conditions in participants' ability to recognize their partners' emotional state.

We also examined the relationship between recognizing partners' self-reported personality traits and avatar appearance. The results were highly similar to those for emotion recognition. However, as with emotional state ratings, these ratings did not differ by condition (all p's larger than 0.100) and there was no difference in trait consensus by condition. For detailed results, please see **Appendix B**.

We summarize these findings as follows. If we assume that self-report is a reasonable "ground truth" for emotional states, then participants were able to identify each other's emotional states at a rate higher than chance. This ability was not significantly affected by the appearance of the avatars in which participants were embodied.

3.2 Emotional Recognition Accuracy and User Experience/Presence

A linear mixed-effect model was used to test whether adaptation/ immersion, involvement, visual quality, audio quality and distraction predict how well participants are able to tell their partners' emotional states. The pair ID is included as a random effect to account for the non-independence of the two partner's ratings.

We found no significant effect of adaptation/immersion ($F = 1.130$, $p = 0.290$) or distraction ($F = 0.166$, $p = 0.684$), involvement ($F = 0.251$, $p = 0.617$), visual quality ($F = 0.095$, $p = 0.758$), audio quality ($F = 0.134$, $p = 0.715$). This result indicates that the participants' experience of a virtual environment does not impact how they interpret their conversational partners' emotional states.

Additionally, we checked whether there was a significant effect of participants' self-presence measures to predict how their conversational partners predict their emotional states. There was no significant effect of self-presence on their partners' interpretation of their emotional states ($F = 0.566$, $p = 0.453$).

3.3 Social Closeness and Social Presence

We found a marginally significant effect of social closeness on how well the participants predicting their conversational partners' emotional states ($F = 3.438$, $p = 0.066$). On the pair level, there was a significant positive correlation between social closeness and how well participants predict others emotional states ($S = 39{,}882$, $p = 0.00197$, $rho = 0.359$).

There was no significant effect of social presence on participants' prediction of their conversational partners' emotional states ($F = 0.264$, $p = 0.609$). On the pair level, there was not a significant correlation between social closeness and their prediction of their conversational partners' emotional states ($S = 52{,}246$, $p = 0.180$, $rho = 0.160$).

3.4 Movement Measures

The lack of significant differences between avatar conditions in participants' perception of each other's emotional states implies that the mere appearance of the avatars did not have an effect on emotion recognition. There are several potential explanations for this. Participants may have focused on the voice, or words, of their partners to get information about their emotional state and other affective information, rather than taking cues from their partner's non-verbal behavior as they would do during face to face interactions. This aligns with previous work in which participants reported using tone of voice as a primary cue for emotional state (Sun et al., 2019), and also with the finding that audio quality predicted emotional recognition accuracy. While we find this explanation plausible, it is also true that both the cube avatar and humanoid avatar conditions were designed to convey similar information about participants' gestures and postures, since in both, participant's head position and hand positions were rendered in the avatars. So, participants could have been using nonverbal behavior similarly in both conditions. This hypothesis is supported by the idea that visual quality also predicted emotion recognition. In fact, other research has found that even when participants report attending primarily to voice they are still influenced by nonverbal behavior (Garau et al., 2003). Thus, we also wanted to explore whether nonverbal cues that would be observable in these avatar conditions could be related to participants' self-reported states of mind, or their partners' estimations of their states of mind.

In order to do this, we selected two nonverbal behaviors to explore. One behavior was on the pair level: proximity, or the distance between the two participants in a pair. The second behavior was on the individual level: expansiveness of gesture. These were the only measures that we generated from movement data in this exploratory analysis.

3.4.1 Proximity

We selected the measure of *proximity* as a between-pairs measure of rapport, following Tickle-Degnen and Rosenthal (1987). To create the proximity measure, for each pair of participants, we calculated the Euclidean distance between two participants' tracked distance between their head-mounted displays using the X and Z positions. We excluded the y position from our measure because they position shows participants' height, which could introduce noise in the measure due to individual differences in height. We then took the average of the distance between their heads over the entire interaction for each pair.

3.4.2 Expansive Gesture

We selected the measure of expansiveness as an individual-level measure that has been used in the literature as an indicator of extraversion in humans (Campbell and Rushton, 1978; Gallaher, 1992; Argyle, 2013) as well as in agent-avatars to express extraversion (André et al., 2000; Pelachaud, 2009). We operationalized expansiveness of gesture by how far apart an individual participant held their hands. To create the expansive gesture measure, for each participant, we calculated the Euclidean distance between their left and right hands using the X, Y, and Z position. Then the distance of the hand movement over the entire interaction was averaged for each participant to create an open gesture measure for each individual. As described in Sun et al. (2019), we filtered out eleven pairs of participants due to either left hand or right hand data missing due to the technical tracking issue before calculating the measure of expansiveness.

If either of these movement measures could be linked to participants' self-reported states of mind, or their estimations of their partners' states of mind, then this would support the possibility that participants were still using nonverbal behavior as information. In order to limit our analyses, these were the only two nonverbal behaviors we explored.

3.4.3 Emotional States and Proximity

We first explored proximity, measured by the distance between participants' heads. Using a *t*-test, we found no significant difference (t (65.324) = 0.491, $p = 0.625$) in proximity between pairs in the humanoid avatar condition ($M = 2.516$, $SD = 0.643$) and the cube condition ($M = 2.433$, $SD = 0.794$).

First, we tested whether the distance between pair members correlated to pairs' average ratings of their own emotional states. Using a Shapiro-Wilk normality test, we found that the proximity measure was normally distributed ($W = 0.970$, $p = 0.065$). Because proximity is a measure taken at the pair level of analysis, we combined participants' emotional state self-reports to get a joint measure of emotional state. We used a Pearson's r correlation test and found that there was a negative significant relationship between participants' joint ratings of how tense they felt and the proximity between their two heads (r (73) = −0.344, $p = 0.003$). In other words, the closer participants were standing to each other, the more tense emotions they reported experiencing. We also found a positive significant correlation between participants' energetic emotion and proximity (r (74) = 0.262, $p = 0.022$). The more participants felt energetic about their own emotional states, the larger their interpersonal distance. However, there was not a significant correlation between hedonic emotion and proximity (r (74) = − 0.148, $p = 0.203$). **Figure 3** shows the correlation between proximity and group self-ratings on emotional states.

3.4.4 Proximity and Social Closeness

We next used a Pearson's r correlation test to check whether proximity is related to social closeness. We found that there was a marginally significant positive correlation between proximity and social closeness (r (74) = 0.202, $p = 0.080$). Surprisingly, the higher distance between the two participants, the *more* social closeness on average that they reported.

In order to further explore this result, we used a linear model examining the interaction between avatar appearance and proximity. In this model, there was no significant interaction, nor any significant main effect of these two variables on social closeness (all p's larger than 0.200). In other words, proximity predicted social closeness ratings no matter what kind of avatar participants used.

3.4.5 Extraversion, Emotional States and Expansive Gesture

Second, we investigated the relationship between individuals' gestures; their self-reported emotional states, and their partners' ratings. To do so we used the measure of expansiveness of gesture reflected by how far apart participants held their hands. In the case of the humanoid avatar, this would have been reflected by the distance between the avatar's hands; in the case of the cube avatar, this would have been reflected by the width of the cube which would have grown wider as participants moved their hands apart. Because some participants' left or right hand data were missing, we dropped 11 participant pairs from our analysis, leaving 65 pairs of participants.

We used the lme4 package in R to test linear mixed effect models. In order to control for differences in participant size, which could also have influenced the distance between hands, we used participant height (operationalized by the mean of the Y-axis position of the head) as a fixed effect in all models. We used pair ID as a random effect to account for the non-independence of the two partner's self-ratings.

We next examined whether participants' self-ratings of extraversion were related to the expansiveness of their gestures. We again included height as a fixed effect and pair ID as a random effect. We found a non-significant difference between expansiveness of gesture and participant's self-reported extraversion ratings, such that participants with higher ratings of extraversion had a slightly greater distance between their hands on average ($F = 1.663$, $p = 0.200$). As the literature would predict, we did not find a significant difference in the rest of the participants' self-rating of personality traits when they had different open gestures (all p's larger than 0.150).

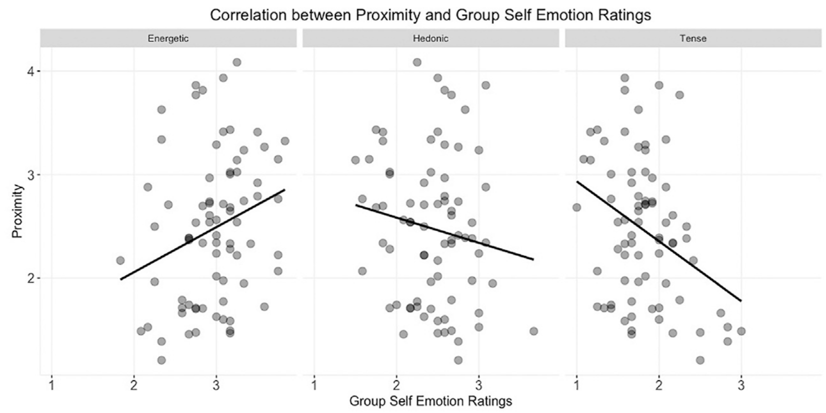

FIGURE 3 | Correlation between proximity and group self-ratings on emotional states.

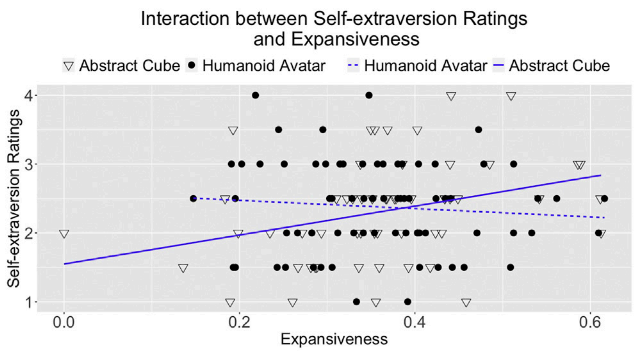

FIGURE 4 | Interaction between self-extraversion ratings and expansiveness.

Finally, we examined whether there was an interaction between avatar appearance and expansiveness of gesture predicting self-ratings of extraversion using a linear mixed-effects model with the lmer function from the lme4 package in R. Appearance and extraversion were used as fixed effects, and pair ID as a random effect in the model. We found a significant interaction between appearance and expansiveness of gesture, such that participants in the cube condition who self-rated themselves as being more extraverted also had more expansive gestures ($F = 6.013$, $p = 0.016$) (see **Figure 4**). We also found a main effect of appearance, such that participants in the cube condition had more expansive gestures overall. This aligns with our observations of participants in the cube condition; when participants first saw their new appearance in the mirror scene, many were intrigued by their ability to make the cube grow and shrink by moving their arms and spent some time playing with this ability, which may have made this gesture more salient.

4 DISCUSSION

In this paper, we explored whether differences in avatar appearance led to differences in participants' perception of their conversational partners' personality or emotional states. Participants' ratings of their partners' emotional states agreed with the partners' self-ratings at a rate significantly higher than chance across both conditions. However, we did not find any significant differences in emotion perception between avatar appearance conditions. We propose two possible explanations for this. First, the gestural and postural information participants received in both conditions might have been equivalently informative. Alternatively, because some important parts of interpersonal communication were missing from virtual reality, participants may not have relied on their partner's movements at all. For example, there was no eye contact, lip sync or facial expressions rendered for participants' avatars in either condition, which are all important information streams that could aid in emotion recognition. In this case, in both conditions, participants might rely primarily on the voice and words of their conversational partner.

To further investigate this, we looked at nonverbal cues that could be communicated through these avatars: proximity and expansiveness of gesture, and whether these clues could be linked to participants' self report of their own or their partner's states of mind. Participants' joint self ratings on energetic emotion were positively correlated with their proximity with each other; participants who reported higher levels of energetic emotion stood further apart. Tense emotion was negatively correlated with proximity; participants who reported higher joint levels of tension stood closer together. Surprisingly, social closeness also correlated *positively* with proximity: participant pairs who expressed higher levels of social closeness stood far apart.

In order to better understand whether this last finding might be a false positive, we explored whether there was an interaction effect of the avatar appearance and open gesture on proximity. Because open gestures caused the cubes to grow in all three dimensions, this may have made their avatar appear closer, participants who were making open gestures may have increased the distance between themselves and their partners to maintain an appropriate interpersonal space. However, we did not find a significant interaction effect: the interaction between avatar appearance no S and the open gestures did not impact the proximity in a statistically significant way. Thus, we are left without a good explanation for the unexpected positive relationship between the distance between participants and social closeness ratings. However, this result does point to the idea that changes in avatar embodiment may have unexpected effects both on emergent nonverbal behavior, and also how people perceive and interpret newly emergent nonverbal behaviors.

Expansiveness of gesture was predictive of self-ratings of extraversion, but only in the cube condition. Participants did not appear to change their ratings of their partner's extraversion according to expansiveness in either condition.

We interpret these exploratory findings as supporting the possibility that even though participants may have primarily used audio channels to determine their partners' emotional states, emergent nonverbal behavior, which may or may not have been informative to their partners, was reflected in their avatar gestures, and this visual information may have also aided emotion recognition. Interestingly, some behavior, like expansiveness of gesture, may have been more salient in the abstract conditions. Further work is necessary to confirm these exploratory findings, and also to determine whether such visually apparent nonverbal behavior is eventually used by conversational partners to aid in the interpretation of states of mind. Even if this is possible, it may take time for participants to learn to extrapolate from their own avatar gestures to interpret those of others.

Notably, there were few relationships between conventional measures of presence and emotion recognition. For this reason, we argue that measures of emotion recognition may be an important and overlooked indicator of usability in social virtual environments, especially in those where the avatar may not closely resemble a human form.

4.1 Limitations

There were several limitations that could be improved in future studies. First, while we have limited our exploratory analyses, and we report all of the analyses we did run, further confirmatory experiments are necessary to build on these findings.

When considering the variable of proximity or interpersonal distance, there are some potential confounds. For example, in the cube condition, if people expand their arms, the cubes will enlarge in all directions. Although the expansion of the cube was meant to resemble the way humanoid avatars would take up more space when their arms were extended, this was not a perfect match because humans would generally extend their arms more in the X axis. When the cube avatars expanded, they expanded in X axis but also in the Y and the Z axes. The Z axis in particular could also give the appearance of increased proximity. Thus, future work that specifically examines proximity should use more precisely designed comparison conditions.

In theory, emotion could be perceived even with the minimal three points available through tracked avatar movements. However, our exploratory work probably does not show the full picture of people's emotional states and personality traits. As tracking improves in consumer systems, future work can examine movement data more granularly.

4.2 Next Steps

In this study we used participants' self-reported emotional states as the "ground truth" of emotional scores. Future work could cross-validate participants' self-reported emotional state and psychological trait scores; for example, by running prosodic analysis on participants' voice recordings as the "ground truth".

Future work could also seek other ways to create virtual humanoid avatars to include more nonverbal features. Some current virtual environments render gaze or mouth movements in social interactions. In some social VR platforms such as Facebook spaces (Facebook, 2019), users can create different emotions by using their hand controllers. All of these nonverbal features could be represented abstractly, to further examine whether this transformed nonverbal behavior is being used to inform emotional state and psychological trait perception.

Finally, our movement measures were intentionally kept simple, and there are many other interesting ways to explore movement trends over time. For example, we could use time series (McCleary et al., 1980, Wei, 2006) to understand how people's proximity and gestures change over time and further explore whether these movement are predictive of people's emotional states and personality traits.

5 CONCLUSION

This study examined our final pre-registered research question on the relationship between emotion recognition, individual and dyadic measures of the proximity and openness of gesture, in the context of humanoid and abstract (cube) avatar appearance. We found no difference in emotional state and personality trait recognition between two different avatar appearances. However, recognition was significantly higher than chance in both instances—people were able to perceive each other's emotional states and personality traits, even when inhabiting an abstract cube-shaped avatar. To help elucidate this result, we explored how emotion correlated to proximity and expansiveness of gesture, and found significant correlations between proximity and emotional states as well as certain personality traits.

Whether emotional and psychological perception was aided by the nonverbal behavior available in both avatar conditions, or was purely dependent on other cues, remains unknown. Further investigation is needed to understand what information streams people perceived from their partners' avatar representations, and how those perceptions influenced emotion recognition in virtual reality.

DATA AVAILABILITY STATEMENT

We have packaged our data for sharing on a permanent server through Cornell Center for Social Sciences. Because movement data can provide sensitive information about individuals (Bailenson, 2018), we will require an IRB before releasing movement data. Our movement data has been packaged with our processing code so that researchers can exactly replicate our findings or run their own analyses.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Cornell University Institutional Review Board. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YS and ASW performed conceptualization, data curation, investigation, methodology, visualization, writing the original draft preparation and reviewing and editing. YS also performed data analysis. ASW also performed supervision.

ACKNOWLEDGMENTS

The authors would like to thank all the research assistants who helped to conduct the experiment and code the qualitative data in the study: Sarah Chekfa, Lilian Garrido, Seonghee Lee, T. Milos Kartalijia, Byungdoo Kim, Hye Jin Kwon, Kristi Lin, Swati Pandita, Juliana Silva, Rachel Song, Yutong Wu, Keun Hee Youk, Yutong Zhou, Joshua Zhu, and Colton Zuvich. We thank Meredith Keller for her help in the pilot study, and Katy Miller for her assistance in proofreading the manuscript. We appreciate Omar Shaikh's work in implementing the software and programming for the original project. We also thank Stephen Parry from the Cornell Statistical Consulting Unit for providing statistical consulting, and Florio Arguillas, Jacob Grippin and Daniel Alexander from Cornell Center for Social Sciences for helping validate the code and archiving the data. We also thank Cornell University and the Ithaca community for participating in this experiment.

REFERENCES

André, E., Klesen, M., Gebhard, P., Allen, S., Rist, T., et al. (2000). "Exploiting Models of Personality and Emotions to Control the Behavior of Animated Interactive Agents," in Fourth International Conference on Autonomous Agents, Barcelona, Spain, June 3-7, 2000.

Argyle, M. (2013). *Bodily Communication*. Routledge.

Astrid, M., Krämer, N. C., and Gratch, J. (2010). "How Our Personality Shapes Our Interactions with Virtual Characters-Implications for Research and Development," in International Conference on Intelligent Virtual Agents, Philadelphia, PA, September 20-22, 2010 (Springer), 208–221.

Atkinson, A. P., Dittrich, W. H., Gemmell, A. J., and Young, A. W. (2004). Emotion Perception from Dynamic and Static Body Expressions in point-light and Full-Light Displays. *Perception* 33, 717–746. doi:10.1068/p5096

Bailenson, J. N., and Yee, N. (2005). Digital Chameleons: Automatic Assimilation of Nonverbal Gestures in Immersive Virtual Environments. *Psychol. Sci.* 16, 814–819. doi:10.1111/j.1467-9280.2005.01619.x

Bailenson, J. N., Yee, N., Merget, D., and Schroeder, R. (2006). The Effect of Behavioral Realism and Form Realism of Real-Time Avatar Faces on Verbal Disclosure, Nonverbal Disclosure, Emotion Recognition, and Copresence in Dyadic Interaction. *Presence Teleop.Virtual Environ.* 15, 359–372. doi:10.1162/pres.15.4.359

Bailenson, J. (2018). Protecting Nonverbal Data Tracked in Virtual Reality. *JAMA Pediatr.* 172, 905–906. doi:10.1001/jamapediatrics.2018.1909

Banse, R., and Scherer, K. R. (1996). Acoustic Profiles in Vocal Emotion Expression. *J. Personal. Soc. Psychol.* 70, 614–636. doi:10.1037/0022-3514.70.3.614

Borkenau, P., Mauer, N., Riemann, R., Spinath, F. M., and Angleitner, A. (2004). Thin Slices of Behavior as Cues of Personality and Intelligence. *J. Personal. Soc. Psychol.* 86, 599–614. doi:10.1037/0022-3514.86.4.599

Campbell, A., and Rushton, J. P. (1978). Bodily Communication and Personality. *Br. J. Soc. Clin. Psychol.* 17, 31–36. doi:10.1111/j.2044-8260.1978.tb00893.x

Clarke, T. J., Bradshaw, M. F., Field, D. T., Hampson, S. E., and Rose, D. (2005). The Perception of Emotion from Body Movement in point-light Displays of Interpersonal Dialogue. *Perception* 34, 1171–1180. doi:10.1068/p5203

Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., et al. (2001). Emotion Recognition in Human-Computer Interaction. *IEEE Signal. Process. Mag.* 18, 32–80. doi:10.1109/79.911197

Dael, N., Goudbeek, M., and Scherer, K. R. (2013). Perceived Gesture Dynamics in Nonverbal Expression of Emotion. *Perception* 42, 642–657. doi:10.1068/p7364

Dael, N., Mortillaro, M., and Scherer, K. R. (2012). Emotion Expression in Body Action and Posture. *Emotion* 12, 1085. doi:10.1037/a0025737

De Gelder, B., and Vroomen, J. (2000). The Perception of Emotions by Ear and by Eye. *Cogn. Emot.* 14, 289–311. doi:10.1080/026999300378824

De Meijer, M. (1989). The Contribution of General Features of Body Movement to the Attribution of Emotions. *J. Nonverbal Behav.* 13, 247–268. doi:10.1007/bf00990296

Dodds, T., Mohler, B., and Bülthoff, H. (2010). "A Communication Task in Hmd Virtual Environments: Speaker and Listener Movement Improves Communication," in 23rd Annual Conference on Computer Animation and Social Agents (CASA), Saint-Malo, France, May 31-June 2, 2010, 1–4.

Doyle, C. M., and Lindquist, K. A. (2017). "Language and Emotion: Hypotheses on the Constructed Nature of Emotion Perception," in *The Science of Facial Expression* (Oxford University Press).

Ehrsson, H. H., Holmes, N. P., and Passingham, R. E. (2005). Touching a Rubber Hand: Feeling of Body Ownership Is Associated with Activity in Multisensory Brain Areas. *J. Neurosci.* 25, 10564–10573. doi:10.1523/jneurosci.0800-05.2005

Enea, V., and Iancu, S. (2016). Processing Emotional Body Expressions: State-Of-The-Art. *Soc. Neurosci.* 11, 495–506. doi:10.1080/17470919.2015.1114020

Facebook (2019). Facebook Spaces. Website. [Dataset]. Available at: https://www.facebook.com/spaces (Accessed September 19, 2019).

Fong, K., and Mar, R. A. (2015). What Does My Avatar Say about Me? Inferring Personality from Avatars. *Personal. Soc. Psychol. Bull.* 41, 237–249. doi:10.1177/0146167214562761

Gallaher, P. E. (1992). Individual Differences in Nonverbal Behavior: Dimensions of Style. *J. Personal. Soc. Psychol.* 63, 133. doi:10.1037/0022-3514.63.1.133

Garau, M., Slater, M., Vinayagamoorthy, V., Brogni, A., Steed, A., and Sasse, M. A. (2003). "The Impact of Avatar Realism and Eye Gaze Control on Perceived Quality of Communication in a Shared Immersive Virtual Environment," in Proceedings of the SIGCHI conference on Human factors in computing systems (ACM), Ft. Lauderdale, FL, April 5-10, 2003, 529–536. doi:10.1145/642611.642703

Garau, M. (2003). "The Impact of Avatar Fidelity on Social Interaction in Virtual Environments,". Technical Report (University College London).

Gosling, S. D., Rentfrow, P. J., and Swann, W. B., Jr (2003). A Very Brief Measure of the Big-Five Personality Domains. *J. Res. Personal.* 37, 504–528. doi:10.1016/s0092-6566(03)00046-1

Gunes, H., and Piccardi, M. (2007). Bi-modal Emotion Recognition from Expressive Face and Body Gestures. *J. Netw. Comp. Appl.* 30, 1334–1345. doi:10.1016/j.jnca.2006.09.007

Hertenstein, M. J., Holmes, R., McCullough, M., and Keltner, D. (2009). The Communication of Emotion via Touch. *Emotion* 9, 566. doi:10.1037/a0016108

Johansson, G. (1973). Visual Perception of Biological Motion and a Model for its Analysis. *Percept. psychophys.* 14, 201–211. doi:10.3758/bf03212378

Kobylinski, P., Pochwatko, G., and Biele, C. (2019). "Vr Experience from Data Science point of View: How to Measure Inter-subject Dependence in Visual Attention and Spatial Behavior," in International Conference on Intelligent

Human Systems Integration, San Diego, February 7-10, 2019 (Springer), 393–399. doi:10.1007/978-3-030-11051-2_60

Koppensteiner, M. (2013). Motion Cues that Make an Impression: Predicting Perceived Personality by Minimal Motion Information. *J. Exp. Soc. Psychol.* 49, 1137–1143. doi:10.1016/j.jesp.2013.08.002

Kozlowski, L. T., and Cutting, J. E. (1977). Recognizing the Sex of a walker from a Dynamic point-light Display. *Percept. psychophys.* 21, 575–580. doi:10.3758/bf03198740

Latoschik, M. E., Roth, D., Gall, D., Achenbach, J., Waltemate, T., and Botsch, M. (2017). "The Effect of Avatar Realism in Immersive Social Virtual Realities," in Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology (ACM), 39. doi:10.1145/3139131.3139156

Li, C., Androulakaki, T., Gao, A. Y., Yang, F., Saikia, H., Peters, C., et al. (2018). "Effects of Posture and Embodiment on Social Distance in Human-Agent Interaction in Mixed Reality," in Proceedings of the 18th International Conference on Intelligent Virtual Agents (ACM), 191–196. doi:10.1145/3267851.3267870

Lorey, B., Kaletsch, M., Pilgramm, S., Bischoff, M., Kindermann, S., Sauerbier, I., et al. (2012). Confidence in Emotion Perception in point-light Displays Varies with the Ability to Perceive Own Emotions. *PLoS One* 7, e42169. doi:10.1371/journal.pone.0042169

Lugrin, J.-L., Latt, J., and Latoschik, M. E. (2015). "Avatar Anthropomorphism and Illusion of Body Ownership in Vr," in 2015 IEEE Virtual Reality (VR) (IEEE), France, March 23-27, 2015, 229–230. doi:10.1109/vr.2015.7223379

Matthews, G., Jones, D. M., and Chamberlain, A. G. (1990). Refining the Measurement of Mood: The Uwist Mood Adjective Checklist. *Br. J. Psychol.* 81, 17–42. doi:10.1111/j.2044-8295.1990.tb02343.x

McCleary, R., Hay, R. A., Meidinger, E. E., and McDowall, D. (1980). *Applied Time Series Analysis for the Social Sciences*. Beverly Hills, CA: Sage Publications.

McVeigh-Schultz, J., Márquez Segura, E., Merrill, N., and Isbister, K. (2018). "What's it Mean to Be Social in Vr?: Mapping the Social Vr Design Ecology," in Proceedings of the 2018 ACM Conference Companion Publication on Designing Interactive Systems (ACM), Hong Kong, China, June 9-13, 2018, 289–294.

Missana, M., Atkinson, A. P., and Grossmann, T. (2015). Tuning the Developing Brain to Emotional Body Expressions. *Develop. Sci.* 18, 243–253. doi:10.1111/desc.12209

Montagne, B., Kessels, R. P., De Haan, E. H., and Perrett, D. I. (2007). The Emotion Recognition Task: A Paradigm to Measure the Perception of Facial Emotional Expressions at Different Intensities. *Perceptual Mot. skills* 104, 589–598. doi:10.2466/pms.104.2.589-598

Mori, M. (1970). Bukimi No Tani [the Uncanny valley]. *Energy* 7, 33–35.

Mousas, C., Anastasiou, D., and Spantidi, O. (2018). The Effects of Appearance and Motion of Virtual Characters on Emotional Reactivity. *Comput. Hum. Behav.* 86, 99–108. doi:10.1016/j.chb.2018.04.036

Nowak, K. L., and Fox, J. (2018). Avatars and Computer-Mediated Communication: a Review of the Definitions, Uses, and Effects of Digital Representations. *Rev. Commun. Res.* 6, 30–53. doi:10.12840/issn.2255-4165.2018.06.01.015

Nowak, K. L., and Rauh, C. (2005). The Influence of the Avatar on Online Perceptions of Anthropomorphism, Androgyny, Credibility, Homophily, and Attraction. *J. Comp.Mediated Commun.* 11, 153–178. doi:10.1111/j.1083-6101.2006.tb00308.x

Pelachaud, C. (2009). Studies on Gesture Expressivity for a Virtual Agent. *Speech Commun.* 51, 630–639. doi:10.1016/j.specom.2008.04.009

Pertaub, D.-P., Slater, M., and Barker, C. (2002). An experiment on Public Speaking Anxiety in Response to Three Different Types of Virtual Audience. *Presence Teleop. Virtual Environ.* 11, 68–78. doi:10.1162/105474602317343668

Roth, D., Lugrin, J.-L., Galakhov, D., Hofmann, A., Bente, G., Latoschik, M. E., et al. (2016). "Avatar Realism and Social Interaction Quality in Virtual Reality," in 2016 IEEE Virtual Reality (VR) (IEEE), Greenville, SC, March 19-23, 2016, 277–278. doi:10.1109/vr.2016.7504761

Schirmer, A., and Adolphs, R. (2017). Emotion Perception from Face, Voice, and Touch: Comparisons and Convergence. *Trends Cogn. Sci.* 21, 216–228. doi:10.1016/j.tics.2017.01.001

Shaikh, O., Sun, Y., and Won, A. S. (2018). "Movement Visualizer for Networked Virtual Reality Platforms," in 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR) (IEEE), Tuebingen/Reutlingen, Germany, March 18-22, 2018, 681–682. doi:10.1109/vr.2018.8446398

Shin, M., Kim, S. J., and Biocca, F. (2019). The Uncanny valley: No Need for Any Further Judgments when an Avatar Looks Eerie. *Comput. Hum. Behav.* 94, 100–109. doi:10.1016/j.chb.2019.01.016

Sun, Y., Shaikh, O., and Won, A. S. (2019). Nonverbal Synchrony in Virtual Reality. *PloS one* 14, e0221803. doi:10.1371/journal.pone.0221803

Tarr, B., Slater, M., and Cohen, E. (2018). Synchrony and Social Connection in Immersive Virtual Reality. *Scientific Rep.* 8, 1–8. doi:10.1038/s41598-018-21765-4

Thomas, J., Azmandian, M., Grunwald, S., Le, D., Krum, D. M., Kang, S.-H., et al. (2017). "Effects of Personalized Avatar Texture Fidelity on Identity Recognition in Virtual Reality," in ICAT-EGVE, Adelaide, Australia, November 22-24, 2017, 97–100.

Tickle-Degnen, L., and Rosenthal, R. (1987). "Group Rapport and Nonverbal Behavior," in *Group Processes and Intergroup Relations* (Sage Publications, Inc.).

Tolani, D., Goswami, A., and Badler, N. I. (2000). Real-time Inverse Kinematics Techniques for Anthropomorphic Limbs. *Graph. Models* 62, 353–388. doi:10.1006/gmod.2000.0528

Waltemate, T., Gall, D., Roth, D., Botsch, M., and Latoschik, M. E. (2018). The Impact of Avatar Personalization and Immersion on Virtual Body Ownership, Presence, and Emotional Response. *IEEE Trans. Vis. Comput. Graph.* 24, 1643–1652. doi:10.1109/tvcg.2018.2794629

Wei, W. W. (2006). "Time Series Analysis," in *The Oxford Handbook of Quantitative Methods in Psychology*, 2.

Witmer, B. G., Jerome, C. J., and Singer, M. J. (2005). The Factor Structure of the Presence Questionnaire. *Presence Teleop. Virtual Environ.* 14, 298–312. doi:10.1162/105474605323384654

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

Won, A. S., Bailenson, J., Lee, J., and Lanier, J. (2015). Homuncular Flexibility in Virtual Reality. *J. Comput. Mediated Commun.* 20, 241–259. doi:10.1111/jcc4.12107

Won, A. S., Shriram, K., and Tamir, D. I. (2018). Social Distance Increases Perceived Physical Distance. *Soc. Psychol. Personal. Sci.* 9, 372–380. doi:10.1177/1948550617707017

Yee, N., and Bailenson, J. (2007). The proteus Effect: The Effect of Transformed Self-Representation on Behavior. *Hum. Commun. Res.* 33, 271–290. doi:10.1111/j.1468-2958.2007.00299.x

Yee, N., Ducheneaut, N., and Ellis, J. (2008). The Tyranny of Embodiment. *Artifact: J. Des. Pract.* 2, 88–93. doi:10.1080/17493460903020398

Yee, N., Harris, H., Jabon, M., and Bailenson, J. N. (2011). The Expression of Personality in Virtual Worlds. *Soc. Psychol. Personal. Sci.* 2, 5–12. doi:10.1177/1948550610379056

Zibrek, K., Kokkinara, E., and McDonnell, R. (2018). The Effect of Realistic Appearance of Virtual Characters in Immersive Environments-Does the Character's Personality Play a Role? *IEEE Trans. Vis. Comput. Graph.* 24, 1681–1690. doi:10.1109/tvcg.2018.2794638

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

APPENDIX A: EMOTIONAL DIFFERENCE SCORES

We replicated the results in ***section 3.1*** with the emotional difference score, which was generated by taking the difference of the participants' ratings of their partners and their partners' rating of themselves, indicating that participants were able to recognize each other's emotional states at a rate significantly different than chance across both conditions ($M = 0.827$, $SD = 0.230$) using a one sample t-test comparing the result with 0 (t (71) = 30.441, $p < 0.001$).

Since the emotional consensus score was not normally distributed ($W = 0.960$, $p = 0.022$) using a Shapiro-Wilk normality test, we used a Wilcoxon rank sum test, finding that there were no significant differences in the emotion consensus scores ($W = 716$, $p = 0.415$) whether participants were represented by humanoid avatars or cubes. In other words, we did not see a difference between conditions in participants' ability to recognize their partners' emotional states.

APPENDIX B: TRAIT RECOGNITION AND AVATAR APPEARANCE

Similar to emotion recognition, we explored whether there was a difference in how personality traits were perceived by participants depending on the appearance of the avatar used during in the interaction. In other words, did participants who saw their partners represented by cubes rate their partner's personality as being more open, conscientious, extroverted, agreeable or neurotic compared to participants who saw their partners represented by humanoid avatars?

Using a linear mixed-effect model, we tested whether the avatar's appearance is a predictor of every participant's ratings to their conversational partners' personality traits. We found no significant difference in participants' ratings of their conversational partners' characteristics of openness to new experiences ($F = 0.087$, $p = 0.769$), conscientiousness ($F = 2.121$, $p = 0.147$), extroversion ($F = 0.319$, $p = 0.574$), agreeableness ($F = 2.300$, $p = 0.132$), emotional stability ($F = 1.177$, $p = 0.282$), regardless of the avatar appearance of their conversational partners.

Secondly, we explored whether participants who saw their partners in a humanoid avatar could perceive their partners' personality traits more *accurately* compared to those who saw their partners represented by an animated cube. We tested whether there was a significant difference in the trait consensus scores and trait perception scores when participants were represented by different avatar appearances.

The traits consensus score, which was generated by averaging the correlation between participants' trait ratings of their partners and their partner's rating of themselves, indicated that participants were able to recognize each other's traits at a rate significantly different than chance across both conditions ($M = 0.394$, $SD = 0.265$) using a one sample t-test comparing the result with 0 (t (68) = 12.339, $p < 0.001$).

Since the traits consensus score is not normally distributed ($W = 0.939$, $p = 0.002$) using a Shapiro-Wilk normality test, we used a Wilcoxon rank sum test, finding that there were no significant differences in the traits consensus scores ($W = 675$, $p = 0.253$) whether participants were represented by humanoid avatars or cubes. In other words, we did not see a difference between conditions in participants' ability to recognize their partners' traits.

frontiers in Virtual Reality

HYPOTHESIS AND THEORY
published: 25 October 2021
doi: 10.3389/frvir.2021.694315

Spatial Presence in Mixed Realities–Considerations About the Concept, Measures, Design, and Experiments

Carolin Wienrich[1]*, *Philipp Komma*[1], *Stephanie Vogt*[1] *and Marc E. Latoschik*[2]

[1]Human-Technology-Systems, Julius Maximilian University of Würzburg, Würzburg, Germany, [2]Human-Computer Interaction, Julius Maximilian University of Würzburg, Würzburg, Germany

Plenty of theories, models, measures, and investigations target the understanding of virtual presence, i.e., the sense of presence in immersive Virtual Reality (VR). Other varieties of the so-called eXtended Realities (XR), e.g., Augmented and Mixed Reality (AR and MR) incorporate immersive features to a lesser degree and continuously combine spatial cues from the real physical space and the simulated virtual space. This blurred separation questions the applicability of the accumulated knowledge about the similarities of virtual presence and presence occurring in other varieties of XR, and corresponding outcomes. The present work bridges this gap by analyzing the construct of presence in mixed realities (MR). To achieve this, the following presents (1) a short review of definitions, dimensions, and measurements of presence in VR, and (2) the state of the art views on MR. Additionally, we (3) derived a working definition of MR, extending the Milgram continuum. This definition is based on entities reaching from real to virtual manifestations at one time point. Entities possess different degrees of referential power, determining the selection of the frame of reference. Furthermore, we (4) identified three research desiderata, including research questions about the frame of reference, the corresponding dimension of transportation, and the dimension of realism in MR. Mainly the relationship between the main aspects of virtual presence of immersive VR, i.e., the *place-illusion*, and *the plausibility-illusion*, and of the referential power of MR entities are discussed regarding the concept, measures, and design of presence in MR. Finally, (5) we suggested an experimental setup to reveal the research heuristic behind experiments investigating presence in MR. The present work contributes to the theories and the meaning of and approaches to simulate and measure presence in MR. We hypothesize that research about essential underlying factors determining user experience (UX) in MR simulations and experiences is still in its infancy and hopes this article provides an encouraging starting point to tackle related questions.

Keywords: mixed reality, virtual-reality-continuum, spatial presence, place-illusion, plausibility-illusion, transportation, realism

OPEN ACCESS

Edited by:
*Richard Skarbez,
La Trobe University, Australia*

Reviewed by:
*Thomas W. Schubert,
University of Oslo, Norway
Florian Daiber,
German Research Center for Artificial Intelligence (DFKI), Germany*

***Correspondence:**
*Carolin Wienrich,
carolin.wienrich@uni-wuerzburg.de*

Specialty section:
This article was submitted to Virtual Reality and Human Behaviour, a section of the journal Frontiers in Virtual Reality

Received: *12 April 2021*
Accepted: *28 September 2021*
Published: *25 October 2021*

Citation:
Wienrich C, Komma P, Vogt S and Latoschik ME (2021) Spatial Presence in Mixed Realities–Considerations About the Concept, Measures, Design, and Experiments. Front. Virtual Real. 2:694315. doi: 10.3389/frvir.2021.694315

INTRODUCTION

The construct of presence is strongly linked with user experience in *Virtual Reality* (VR) (e.g., Skarbez et al., 2018). The feeling of being there (i.e., virtual spatial presence) can probably be considered as a so-called *hygiene factor* (Wienrich and Gramlich, 2020). To a certain extent, it might be necessary to allow other VR potentials to become effective. It might be conceptualized similarly to the role of pragmatic quality within the field of user experience (Hassenzahl, Diefenbach, and Göritz, 2010). The emergence of presence is often determined by the allocation of attention to the virtual environment and the occlusion of the physical environment. Wirth et al. (2007) define this allocation/occlusion process as reference setting, using the (virtual or physical) environment as referential cues. Persons have a sense of virtual presence if attention is allocated to immersive factors, and the virtual environment is chosen as the primary reference frame. Under the umbrella of this understanding, plenty of models, measures, and investigations target the understanding of presence in virtual reality (e.g., Slater, 2009). However, virtual reality is only one possible variant of the *mixed reality continuum* (Milgram and Kishino, 1994). Other variants have less immersive features (at least considering the inclusive dimension of the Inclusive-Extensive-Surrounding-Vivid characteristics, short IESV-characteristics (Skarbez and Whitton, 2017), see below) and continuously interfere with the real physical space. Those variants that allow for fluent transitions between virtual and physical realities might alter the idea of reference setting using different referential cues as the environment (the space). The occlusion of the physical world cannot not be such a crucial criterion inducing a sense of presence beyond full immersive VR. Hence, questions arise about the interplay between virtual and physical referential cues and their consequences for corresponding outcomes (e.g., the sense of transportation, the sense of realism). The present work aims to analyze the construct of presence and the interplay of virtual and physical referential cues in the context of *mixed realities* (MR). To achieve this goal, we present 1) a short review of definitions, dimensions, and measurements of presence in VR, and 2) the state of the art views on MR. Furthermore, we 3) derive a working definition of MR implying new conceptual ideas for the reference setting beyond full immersive VR and environmental cues. In our opinion, in order to discuss spatial presence in MR, it must be assumed that environmental entities are detached as anchor cues. Based on these new ideas, we 4) identify three research desiderata, including research questions about the referential power of entities occurring in MR, the corresponding dimension of transportation (i.e., *place-illusion* in VR), and the dimension of realism (i.e., *plausibility-illusion* in VR) in MR. Finally, we 5) suggest an experimental setup to reveal the research heuristic behind experiments investigating presence in MR. In sum, the current work presents an alternate conceptual idea of reference setting in MR, which raises the question: Does spatial presence in MR refer to the sense of being anywhere (space-related, inside-out) or to the sense of being with something (object-related, outside-in) somewhere? The following presents a research heuristic to investigate the idea and resulting question. Since presence is the most investigated construct evaluating VR experiences, we raise no claim to completeness. Our considerations are limited to the subconstruct of virtual presence and current debates about the spatial component of presence in VR. Nevertheless, we hope to encourage a discussion of the meaning, measurement, and design of spatial presence in MR. We suggest that empirical studies that are described as paradigmatic convey the questions into a coherent set of assumptions, measurements, and useful design suggestions for further research.

THEORETICAL BACKGROUND

Recent Views on Spatial Presence in Virtual Reality

Definition of Presence in the Sense of Spatial Presence in Virtual Reality

This paper considers the construct of *presence* in the context of technologically mediated realities such as VR, *augmented reality* (AR), *augmented virtuality* (AV), and MR. In general, presence refers to subjective perceptions and feelings occurring in those realities by different immersive factors. Many sub-constructs refer to the term of presence, such as *social presence* (e.g., Lee, 2004), *co-presence* (e.g., Slater, 1999), *story presence* (e.g., Brown et al., 2003), *cognitive presence* (e.g., Nunez and Blake, 2001), *relational presence* (e.g., Maguire and Connaughton, 2006), and *spatial presence* (e.g., Lee et al., 2004). Since considering all sub-constructs is beyond the scope, the present paper focuses on *spatial-related definitions of virtual presence.*

Almost all definitions of spatial presence refer to the spatial context in which the term should be used. Gibson (1966) wording of the experience of presence as " [...] the sense of being in an environment" has been the basis for the definition of spatial presence as "being there" (in Steuer, 1992, *p*.75). Sheridan (1992) referred for the first time to spatial presence - not a real place but a *virtual presence* - as "feeling like you are present in the environment generated by the computer" (Sheridan, 1992). Minsky (1980) directed the discussion to the subjective sensation and termed spatial presence as "the sense of being there," referring to the most common use today (Skarbez et al., 2017, 2018). Spatial presence has also often been related to the sense of *transportation* (Lombard and Ditton, 1997). The researchers distinguish between three different types of transportation. Firstly, when the user is transported to another place, secondly, the transportation of another place and objects to the user, and thirdly, two or more communicators are transported to a shared place (Lombard and Ditton, 1997). In VR, spatial presence has been mainly connected to the first type of transportation, i.e., self-transportation. Diverse authors described the "being there" aspect of spatial presence with different words such as the *illusion of non-mediation* (Lombard and Ditton, 1997) or the *place-illusion* (Slater, 2009). In the current article, those spatial-related definitions of presence in VR are named virtual presence, i.e., the feeling of being in an environment generated by an immersive computer system. In contrast, spatial presence encompasses a wider meaning and refers to the sense of being in an environment, including artificial, semi-artificial, and real.

Further, immersion, or rather immersion factors, defined as objective system factors, influence the *place-illusion. Primarily,* immersion is described by the following characteristics (Slater and Wilbur, 1997; Skarbez et al., 2017):

- Inclusive (I) indicates the extent to which physical reality is shut out.
- Extensive (E) indicates the range of sensory modalities accommodated.
- Surrounding (S) indicates the extent to which this virtual reality is panoramic rather than limited to a narrow field.
- Vivid (V) indicates the resolution, fidelity, and variety of energy simulated within a particular modality.

All variants of MR can differ in the characteristics of extensiveness, surrounding, or vividness. However, only the variant of VR possesses the characteristic of inclusiveness. Consequently, when discussing the relation between immersion and virtual presence, the relationship between inclusiveness and virtual presence is meant.

For more discussion about different uses of the term immersion, see, e.g., (Skarbez et al., 2017; Wienrich and Gramlich, 2020).

The second aspect of spatial presence concerns the plausibility of the experience in VR. While the *suspension of disbelief* describes a general willingness to accept objects or events that are not physically real (Slater and Usoh, 1993), the *plausibility-illusion* refers to " [. . .] the illusion that what is happening is real (even though you know that it is not real)." (Slater, 2009). In this view, virtual presence is defined as *being there plus*–the sense of being in the virtual world and (plus) feeling that the events are plausible within this world (Skarbez et al., 2017). The plausibility is often connected to the perceived action possibilities (Wirth et al., 2007) or the richness of interaction (Schubert, 2009). Skarbez et al. (2017) introduced coherence as a set of reasonable circumstances that influences the plausibility illusion. The authors have seen it as parallel to the role of immersion for the *place-illusion* in VR.

Models of the Emergence of Presence in Virtual Reality

Besides diverse approaches of definition, different models try to answer how the sense of presence emerges. Two-pole models assume that a VR application user always feels present in one of two environments, either the real environment or the virtual environment (Biocca, 2003). From the real to the virtual environment, movements on the spectrum are explained by increasing immersive factors of the virtual environment and the amount of attention paid to those factors, such as the field of view or the VR's interactivity. The same is true the other way around, as sudden interruptions, and distractions from the real world can cause *breaks in presence* (BIP; Slater and Steed, 2000). Similarly, Wirth et al. (2007) described a two-level model. Different factors contribute to the construction of a *primary egocentric reference frame* (PERF). Persons have a sense of virtual presence if attention is allocated to immersive factors, and the virtual environment is chosen as the primary reference frame. Further, personal factors such as involvement and the suspension of disbelief impact the attentional shift (i.e., the shift of reference frame) from the real to the virtual world. Wirth et al. (2007) conceived the emergence of virtual presence as a binary sensation, although they, at least theoretically, argued for the possibility of consecutive sensations of virtual presence. Other models added a third pole by integrating the *mental imagery space* (Biocca, 2003). This pole represents fictitious environments created by imagination. According to this model, virtual presence is a continuous state influenced by the position on the three-dimensional spectrum. Thus, some emerging models regard virtual presence sensation as an all-or-nothing principle, others as a continuous state (for a detailed discussion, see Nunez, 2007). Latoschik and Wienrich (2021) proposed an alternative theoretical model describing how XR experiences, including the many variants of presence, emerge. Their model integrates plausibility (Slater, 2009; Skarbez et al., 2017) and coherence (Skarbez et al., 2017) as much more central states or conditions during an XR exposure. They further argue that "there is no plausibility illusion but merely plausibility" with plausibility being defined "as a state or condition during an XR experience that subjectively results from the evaluation of any information processed by [...] sensory, perceptual, and cognitive layers" (Latoschik and Wienrich, 2021). Hence, in their view there are no illusions of the different qualia but just qualia and states. However, in the scope of the current paper we adhere to the widely used illusion terminology in harmonization with the current literature.

Measuring the Dimension of Spatial Presence in Virtual Reality

Different researchers have proposed different operationalizations and related measure methods for presence in VR. However, probably the most common way to capture presence is by *post-experience-questionnaires* (PEQ, also referred to as *post-immersion-questionnaire*). PEQ are self-report questionnaires that are answered following a VR experience (e.g., Insko, 2003; Skarbez et al., 2017). The most commonly used questionnaires based on their citations (Schwind et al., 2019) are the *presence questionnaire* (PQ), the *immersive tendencies questionnaire* (ITQ; Witmer and Singer, 1998), the *SUS presence questionnaires* (SUS; Slater et al., 1998; Usoh et al., 2000, 1999), the *igroup presence questionnaire* (IPQ; Schubert et al., 2001; Schubert, 2003), and the *ITCSense of presence inventory* (ITC-SOPI; Lessiter et al., 2001). Although each questionnaire refers to slightly different scales, three dimensions are essential: immersion, transportation, and realism (Lombard and Ditton, 1997). As stated above, immersion is defined as objective system factors influencing the *place-illusion* (e.g., Skarbez et al., 2017; Wienrich and Gramlich, 2020). Thus, the dimension of transportation and realism is more interesting for the present work.

The Dimension of Transportation

The sense of transportation refers to the crucial dimension of virtual presence (Lombard and Ditton, 1997). All PEQ included transportation questions assessing the sense of "being there"

(Steuer, 1992). Thus, transportation is closely linked to the *place-illusion* in VR. Although Lombard and Ditton (1997) distinguished between three different types of transportation, spatial-related definition presence in VR has been mainly connected to the first type, i.e., self-transportation to another place.

For example, the current version of the SUS by Slater et al. (2000) includes six questions, which can be answered on a 7-point Likert scale. The dimension of transportation is, e.g., measured with the "sense of being in the [. . .] space" (Slater et al., 2000) and with the question, whether the user "think [s] of the [...] space more as images that [he] saw, or more as somewhere that [he] visited." (Slater et al., 2000).

The Dimension of Realism

Following Skarbez and colleagues (Skarbez et al., 2017), realism refers to fidelity, including physical, functional, and psychological sub-categories (see Alexander et al., 2005). It depends on the consistency of the virtual experience. Thus, realism does not describe how well the virtual environment or experience resemble the physical reality, but the effectiveness of the *plausibility-illusion*. When objects, entities, events, or actions make sense in the place accepted as real (i.e., *place-illusion*), users would indicate a high sense of realism. "Coherence can be thought of as a superset of realism or fidelity." (Skarbez et al., 2017, *p*.6). As immersion determines the *place-illusion* and the sense of transportation, coherence refers to a set of reasonable circumstances that influences the *plausibility-illusion* and the sense of realism. For example, the PQ by Witmer and Singer (1998) measures the dimension of realism on behalf of seven questions. e.g., by determining the degree, the user felt "confused or disoriented at the beginning of breaks or at the end of the experimental session" or "how well [the user] could identify sounds" during the experience (Witmer and Singer, 1998). According to the different approach of Latoschik and Wienrich (2021), they extended the significance of coherence. Within their model of XR experiences, coherence activations occur on every level of information processing, including sensory, perceptual, and cognitive, and they elicit all qualities of experiences solely based on the respective cues, e.g., including the place-illusion. However, in the current paper's scope, we adhere to the widely used terminology in harmonization with the current literature.

Summary

In sum, presence in VR is mainly defined by two aspects - the self-transportation to another place (i.e., *place-illusion*, *sense of being there*) and the feeling that events at this place are real (*plausibility-illusion*, sense of *being there plus*). Focussing on technological-mediated experiences, most emerging models suggest two poles of the *place-illusion*–the real environment and the virtual environment (e.g., Slater and Steed, 2000). Others define the poles as two possible manifestations of the *primary egocentric reference frame* (PERF, Wirth et al., 2007) within which events and interactions become plausible. Most authors assume that the sense of being in one of the environments follows an all-or-nothing principle. At the same time, others conceive it as a continuous state with more or less presence in one of the environments (e.g., Biocca, 2003). Measurements focus on the transportation dimension assessing the effectiveness of the *place-illusion*, and the realism dimension assessing the *plausibility-illusion* effectiveness. The present contribution revisited this knowledge corpus for virtual experiences, allowing for a fluent transition between virtual and physical realities, such as many MR headsets with see-through functions or MR applications defined as continuous transitions between the poles (Milgram and Colquhoun, 1999). We particularly question the current view on the all-or nothing principle of reference setting in MR variants where users perceive virtual and physical cues simultaneously. Before discussion of a modified referential cue model, recent views on MR are presented to show why it is important to rethink the concept of presence beyond full immersive VR experiences.

Recent Views on Mixed Realities

There is no unified definition of MR (Speicher and Nebeling, 2019). The most well-known and most cited definition in academic research is that of Milgram and Kishino (1994). The authors formulate MR as "a particular subset of Virtual Reality (VR) related technologies that involve the merging of real and virtual worlds somewhere along the "virtuality continuum" which connects completely real environments to completely virtual ones.". Another key feature of MR environments, according to Milgram and Kishino (1994), was that in "a Mixed Reality environment [...] objects are presented together within a single display". Finally, the authors themselves defined their primary work as: "non-exhaustive examples of existing display systems in which real objects and virtual objects are displayed together.". Milgram and Colquhoun (1999) reiterate the original notion of Milgram and Kishino (1994) by explicitly excluding the extrema on this continuum ("completely real" and "completely virtual") from their MR definition. MR stated as such, therefore only includes environments where real and virtual content is being mixed. Further, Milgram and Colquhoun (1999) extend the conceptualization of MR by showing the *mixed reality combination space*, presenting different combinations of real and virtual content within the same display. They firstly show how different types of hybrid displays may be mapped onto the MR combination space, but also go on to "extend the concept somewhat" by describing a journey along with the combination space within a single application. They never explicitly state such a transition as an MR application, but use it as an example of why the terms "AR" and "AV" cannot always clearly distinguish different states on the virtuality continuum and formulate the need for the broader definition of MR.

Others conceived of MR as even broader by including non-technological mediated phenomena such as realities perceived in dreams or drug experiences. Examples are seen in the work of Hillstead (2017), Mann (2002), and Mann and Nnlf (1994). However, these ideas are out of the scope of this paper because the present considerations focus on technic-mediated perceptions of reality.

More recently, Speicher et al. (2019) analyzed the usage of MR in academia. Through expert interviews and a literature review, they tried to gather different views on the term MR. Researchers used the definition of Milgram and Kishino (1994) most widely. Other conceptualizations included MR either as a synonym for AR, a strong AR, a combination of AR and VR, a type of collaboration, or an alignment of environments. The authors concluded that MR could mean many different things, depending on the context. Further, they recognized the need for researchers to clearly describe their understanding of the term MR within the context of their work. For that purpose, Speicher et al. (2019) described five main MR dimensions: 1) *the number of environments* (one or many), 2) *the number of users* (one or many), 3) *the level of immersion* (not, partially, entirely), 4) *the level of virtuality* (not, partially, entirely), 5) *the degree of interaction* (implicit, explicit). In addition, they introduced *input* and *output* as lower-level dimensions. Although Speicher and colleague's work contributes to the clarity of MR in academic research, transitions (combinations of real and virtual content) are not explicitly discussed. However, *as introduced by* Billinghurst et al. (2001), *transitional interfaces* or the MR headset's *see-through* function cannot be classified with the dimensions. Transitional interfaces allow the interpolation between physical real and virtual environments, and they meet the definition of MR applications as introduced by Milgram and Colquhoun (1999). In addition, the dimensions remain unclear in their descriptions of space and interactions in MR, metaphors being essential for discussions about the meaning of the *place-illusion* and *plausibility-illusion* in MR.

Spatial Presence in Mixed Realities

Only a few scientific publications address the intersection between spatial presence and MR. Wagner et al. (2009), for example, compared three different MR applications. The applications, *MapLense*, a mobile AR system, *TimeWarp*, an augmented reality game, and *MR Tent,* were not classified by the authors. Schaik et al. (2004) examined a collaborative MR application, i.e., *Dessert Rain*. The authors argued that evaluating these MR applications using the standard methods established in VR scenarios (see above) is not necessarily meaningful. Despite the recommendation to use other factors such as ecological and cultural factors, they did not provide any solution to conceive spatial presence in MR.

Moreover, the work of Schaik et al. (2004) and Wagner et al. (2009), despite their scientific contributions, demonstrated, again, a problem in the research of MR applications. There is no unified definition of the term MR, and because of this scientific investigation collected under the term may vary considerably in content and or scope. Also, contributions such as that of Billinghurst et al. (2001) are mostly hidden to researchers because they lack MR as a label. The understanding of the term MR remains unclear, and the investigated MR applications incomplete. For example, to the author's best knowledge, the spatial presence was not the object of research for interfaces allowing the interpolation between physical real and virtual environments as defined by Milgram and Colquhoun (1999).

Outline of Present Contribution

The emergence of virtual presence is often determined by the allocation of attention to the virtual environment and the occlusion of the physical environment. Wirth et al. (2007) define this allocation/occlusion process as reference setting using the (virtual or physical) environment as referential cues. However, virtual reality is only one possible variant of the *mixed reality continuum* (Milgram and Kishino, 1994). Other variants have less immersive features and continuously interfere with the real physical space. Those variants that allow for fluent transitions between virtual and physical realities question the interplay between virtual and physical referential cues and their consequences for corresponding outcomes (e.g., the sense of transportation, the sense of realism). The present work aims to analyze the construct of presence and the interplay of virtual and physical referential cues in the context of MR. In the following, we derive a working definition of MR implying new conceptional ideas for the reference setting beyond full immersive VR and environmental cues. In our opinion, detaching from environmental entities as anchor cues is the essential assumption to discuss the spatial presence in MR. Based on these new ideas, we identify three research desiderata, including research questions about the referential power of entities occurring in MR, the corresponding dimension of transportation (i.e., *place-illusion* in VR), and the dimension of realism (i.e., *plausibility-illusion* in VR) in MR. Finally, we suggest an experimental setup to reveal the research heuristic behind experiments investigating presence in MR. In sum, the current work presents an alternative conceptional idea of reference setting in MR which raises the question: Does spatial presence in MR refer to the sense of being anywhere (space-related, inside-out) or to the sense of being with something (object-related, outside-in) somewhere? It also presents a research heuristic to investigate the idea and resulting question.

WORKING DEFINITION OF MIXED REALITIES

Scope

Similarly to Mann (2002) and Hillstead (2017), the scope of the present work refers to the role of spatial presence concerning their fourth area, i.e., computer-generated realities. The role of spatial presence in the first order, second-order, or mediated realities such as the sense of spatial presence during dreams, reading books, or watching movies are outside of the scope.

Defining the Modified Reality-Virtuality-Continuum

In the following of the present work, the working definition of MR is considered as concrete computer-generated virtual reality traversing the *modified Reality-Virtuality-Continuum* (short mRVC). Similar to the continuum defined by Milgram and Colquhoun (1999), the poles of the mRVC are defined as *natural* or *physical reality* (R) and *virtuality* (V). While the poles R and V represent abstract and theoretical forms of reality, the positions between them refer to concrete forms of

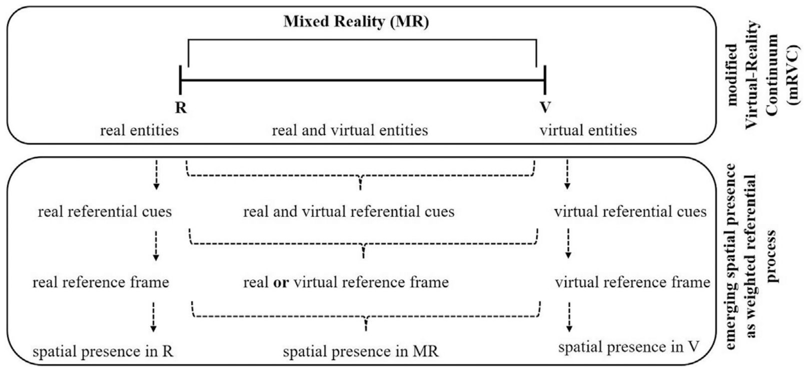

FIGURE 1 | The **upper panel** illustrates the *modified Virtual-Reality-Continuum* (mRVC). The continuum is based on Milgram and Colquhoun (1999) continuum with the addition of real and virtual entities possessing a certain degree of referential power. The **bottom panel** illustrates that entities can become referential cues determining, in turn, the selection of the referential frame and the emergence of presence in MR.

reality illusions (**Figure 1**). Each of these concrete forms of reality illusion defines a mixed reality illusion at one time point. MR must include a merging of real and virtual qualities.

In addition to Milgram and colleague's definitions, firstly, each reality illusion includes entities reaching from real to virtual manifestations at one time point. Entities can refer to spatial environments (i.e., space-based) or concrete objects (i.e., object-based). Secondly, referential cues determine the selection of the reference frame, e.g., allocentric vs egocentric and inside-out vs outside-in. Cues are stemming from the presented entities themselves but also from introduction texts or other framing experiences. We only refer to referential cues stemming from the entities presented as real or virtual manifestations in an MR experience. Thus, spaces can be entities and can serve as cues simultaneously. In previous views, for example, virtual objects represented in virtual spaces refer to VR. Real objects represented in virtual spaces refer to an AV and virtual objects in real space to an AR. Based on the views about virtual presence presented above, the sense of virtual spatial presence has been implicitly anchored in the virtual space. Thus, the virtual spatial environmental entity (space-based) is the cue determining the frame of reference. However, assuming the spatial environmental entity is the cue determining the frame of reference reduces convincibility in MR (**Figure 1**). Why should the virtual environment in AV serve more likely as an anchor cue than the real object user interacting with? Thus, the view presented here overcomes the distinction of environments and objects by introducing entities possessing different degrees of referential power.

The referential power is defined as a weight indicating the probability of each entity or a class of entities (such as the spatial environment) to be selected as a referential cue. These probabilities are not independent of each other but are context-sensitive, and are a result of a given spatial configuration of entities and the location of the user relative to these entities, i.e., her current perspective she has on the entities. For example, an egocentric outside-in view of an entity that does not convey any self-location - imagine a manipulation of a CAD object of an engine part in front of the user - will most likely not elicit a space-based sensation of a *being-there*. In contrast, simple entities resembling floor tiles, geometrically arranged in congruence with a user's perceived floor will most likely elicit a space-based sensation. When multiple entities are perceivable by the user, the question becomes how their referential power compares, how strong the respective cues are in a relation, how congruent they are to each other, and–in the case of the various XR variants defined by the mRVC, the ratio between the real physical and the simulated entities and their relative coverage of a user's field of view.

It follows from the preceding that spatial environmental entities and objects theoretically can serve as referential anchor cues but that the actual effect is dependent on a combination of various conditions as given for a specific scene and entity configuration. In our opinion, detaching from environmental entities as anchor cues is the essential assumption to discuss spatial presence in MR. From this point of view, we discuss in the following subsection how we can operationalize the degree of real or virtual entity manifestations of MR experiences as one possibility to determine objectively the position of the mRVC. Subsequently, we discuss three research desiderata addressing the referential power of entities.

Defining Positions on the mRVC

Although the authors agree with Milgram and Colquhoun's (1999, p. 8) assertion that " [...] determining whether an image should be considered augmented reality or augmented virtuality is also not necessarily a matter of simply summating the respective areas of real and virtual images [. . .]", the mRVC visualized areas defined by the amount of real and virtual picture portions at one time point (**Figure 2**, upper panel). **Figure 2** (lower panel) shows a stepwise interpolation of Varjo demonstration on the mRVC (Varjo, 2019). The mRVC uses a naive percent-visualization of real and virtual picture portions, with 100% referring to reality. A higher portion of cues engendering the perception of real entities leads to a more left-side position on the mRVC. In contrast, cues engendering the perception of virtual entities lead to a more right-side position on the mRVC. Notably, experiences refer to a continuous stream of feelings, thoughts, and actions (Kahneman et al., 1999). Hence,

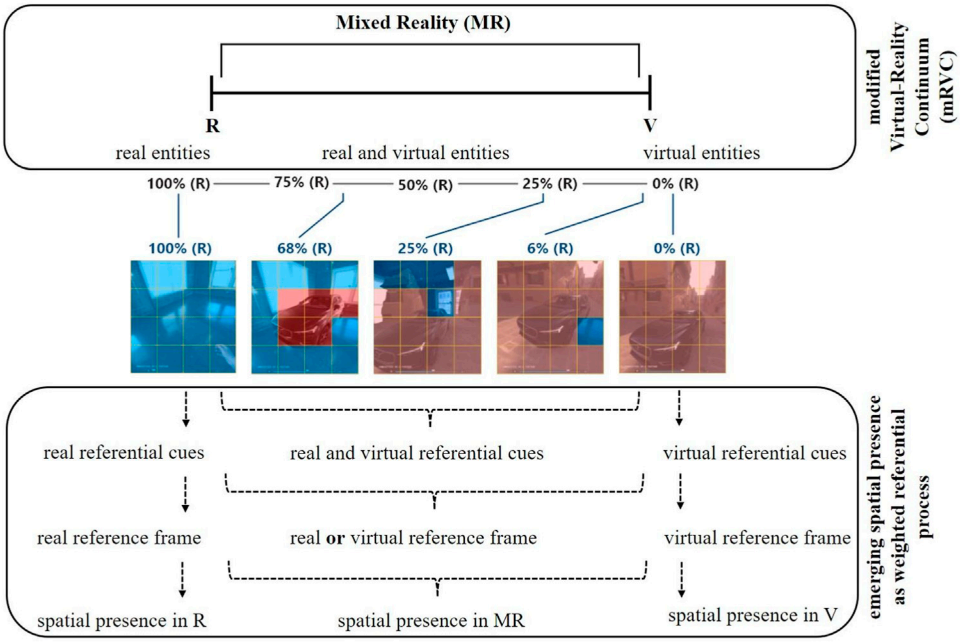

FIGURE 2 | Illustrates the determination of reality (blue) and virtuality (red) portions on the mRVC being useful for experimental setups testing assumptions about the spatial presence in MR.

this position finding does not reflect user's experiences but helps to define and classify reality illusion and thus support systematic testing. Further, such a definition cannot capture subjective attentional allocation to particular areas of interest (e.g., interaction areas). Nevertheless, the possibility to define and classify reality illusions might be one crucial step in enabling a systematic investigation of the entity's referential power and spatial presence in MR.

In the following, we consider the relationship between the main aspects of presence in VR, i.e., the *place-illusion*, and *the plausibility-illusion*, and the entities occurring in the mRVC, i.e., *space-based* and *object-based*, to discuss the concept, measures, and design of spatial presence in MR by considering the referential power of the different entities.

RESEARCH DESIDERATA AND RESEARCH QUESTIONS ABOUT SPATIAL PRESENCE IN MIXED REALITIES

Considering the adaptation of spatial presence for MR, three research desiderata arise concerning: the reference frame (short: RD1), the corresponding dimension of transportation (short: RD2), and the dimension of reality in MR (short: RD3). Each desideratum includes three classes of research questions; questions concerning the construct (short: C-RQ), the ones concerning the measurement (short: M-RQ), and the ones about the design challenges (short: D-RQ). Notably, the desiderata are theoretically derived from knowledge about the spatial presence in VR presented earlier. Further, the desiderata and questions are based on the working definition of MR given above. Of course, other approaches might also be possible. Thus, we raise no claim to completeness but hope to encourage discussion of the meaning of spatial presence in MR.

RD1: Research Desiderata Concerning the Reference Frame in Mixed Realities

More or less directly, considerations about the presence in VR insert the virtual space as an anchor or reference. Often, the real environment serves as a kind of *comparison-reference* that should enable users to report their virtual experience (e.g., in the IPQ the user is asked whether "the virtual world seemed more realistic than the real world.", Schubert et al., 2016). Nevertheless, the virtual and real environments are considered as opponents–only one can be the only (or at least the dominant) reference to judge the experience's plausibility. Others framed the selection of reference as an attentional allocation process. If the virtual environment obtains (more) attention, users feel (more) present in it. In contrast, MR allows for a fluent transition between virtual and physical realities questioning the view of opponent realities. On the other hand, overcoming the distinction between environments and objects (by introducing entities possessing different degrees of referential power) leads to the RD1, including the following research questions:

C-RQ 1.1: Do users have/need a dominant reference frame in MR experience?
C-RQ 1.2: Which *entity* determines the reference frame in MR experience?

M-RQ 1.1: How can it be measured whether users insert or alternate a reference frame during the MR experience?
M-RQ 1.2: How can it be measured which *entity* determines the reference frame during the MR experience?
D-RQ 1.1: Which MR experiences should offer a dominant reference frame, which an alternating one?
D-RQ 1.2: Which design implications result from a dominant or alternating (*entity*) reference frame?

The authors identify three views on the RD1 described in the following.

1) One view to answer the research questions might be the view of Wirth et al. (2007). The authors assumed the *place-illusion*, i.e., the self-positioning in a real or virtual environment, determines the *primary egocentric reference frame* (PERF). Following this assumption, the user's positioning in a *real space* would lead up to the user's perception as part of the real space (*space-PERF = R*). In contrast, positioning the user in a *virtual space* would lead up to the user's perception as part of the virtual space (*space-PERF = V*). However, in mixed realities, besides the self-positioning, the positioning of the *objects* further determines the perception of the illusion. Consequently, the positioning of *objects* in the *real space* would also lead up to the perception that they are part of the real place (left side of the mRVC, see **Figure 1**). In contrast, the positioning of *objects* in the *virtual space* would lead up to the perception that they are part of the virtual place (right side of the mRVC, see **Figure 1**). Thus, one view might be that an MR illusion emerges by integrating *virtual objects* into the *real space* (*space-PERF = R*). Then the user accepts the *virtual objects* as a part of the *real space.* In contrast, another MR illusion emerges by integrating *real objects* into the *virtual space* (*space-PERF = V*). Then the user accepts the *real objects* as a part of the *virtual space.* In this view, *space-based entities* would possess more referential power, become more likely to be referential cues, and determine the *PERF* more likely than *object-based entities.*
2) Another view to answer the research questions is that *the object-based entities* would possess more referential power, become more likely to be referential cues, and determine the *PERF* more likely than *space-based entities.* Then *space-based entities* are only an additional cue that indicates the concrete reality (i.e., position on the mRVC). *Real object-based entities* would then lead up to the user's perception as part of the real place (*object-PERF = R*). Some MR-illusion emerges by integrating *virtual spaces* into the *object-R-PERF.* Then the user accepts the *virtual space* as a part of the *real object-based - place-illusion.* In contrast, *virtual objects* would lead up to the user's perception as part of the virtual place (*object-PERF = V*). The MR illusion then emerges by integrating a *real space* into the *object-V-PERF.* Then the user accepts the *real space* as a part of the *virtual object-based-place-illusion.*

Both views would assume one dominant reference frame, either the quality of *space-based or object-based entity.* The former view would result in similar considerations about MR measures and design as it is in VR. In contrast, the latter view would result in new considerations about measures and design in MR. However, established spatial-related presence measures in VR assess transportation dimension and ask more or less after the degree of being in the experienced VR. In our opinion, transportation plays a unique role in experiencing MR (see 4.2). Instead of referring to the sense of transportation, we would argue for including direct questions about the reference frame or rather about the referential power of entities. One operationalization could be to assess the user's expectations or breaks in expectations relating to the *place-illusion* emerged by a real or virtual *space* or the place spanned by the *object-based entities.* Those questions change the balance when a real and a virtual (*space-based*, *object-based*) PERF would make a huge difference.

It would be essential for design to know if users form a PERF and if the *space-based* or the object-based *entities* determine the reference frame. Then the MR application should include easy to perceive cues indicating clearly to a (*space*, *object*) *R-PERF* or a (*space*, *object*) *V-PERF*. For example, if users form a *space-R-PERF*, objects should cast a shadow in a way that would be plausible under real light conditions.

3) A third view of answering the research questions might be that users alter the PERF continuously or just do not have one. Then probably, the amount of real and virtual picture portions during the experience determine the position on the mRVC. In this view, space could be seen as the sum of its objects. The higher the number of *virtual entities* (including space as one cue), the more likely the user experiences a position on the right side of the mRVC (**Figure 1**). Then users would expect either *real or virtual entity behavior* and would not experience breaks by transitions. Measuring the absence of an experience is challenging (see below). For design, in contrast, this possibility enables much openness.

RD2: Research Desiderata Concerning the Dimension of Transportation in Mixed Realities

The dimension of transportation refers to the *place-illusion* in VR (i.e., the sense of being there, Steuer, 1992). Although Lombard and Ditton (1997) distinguished between three different types of transportation, spatial presence in VR has been mainly connected to the first type, i.e., self-transportation to another place. However, in MR, self-transportation might only be one aspect of transportation (*egocentric inside-out*). The second type, addressing the transportation of another place and objects to the user, probably draws more attention to objects-based entities (*egocentric outside-in*). The sense of spatial presence might also be determined by the feeling of sharing the MR space with virtual or real objects emphasizing the RD2, including the following research questions:

C-RQ 2.1: Does the sense of spatial presence in MR include the first and the second type of transportation?

C-RQ 2.2: Does the sense of transportation in MR refer to a change of reference frame?
M-RQ 2.1: How can the first and the second type of transportation be measured?
M-RQ 2.2: How can the relation of transportation and a change of reference frame be measured?
D-RQ 2.1: If necessary, how can techniques that enable the first or second type of transportation be designed?
D-RQ 2.2: If necessary, how can techniques that support a change of reference frame be designed?
D-RQ 2.3: Are similar transport metaphors as used in VR appropriate in MR?

The authors detected three views on the RD2 described in the following.

1) The established self-transportation probably refers to the views regarding the rest frame described above. In VR, transportation is closely related to a change of space within the VR experience (e.g., from virtual Europe to virtual Africa) or between the real and the virtual space (e.g., at the beginning of an experience when putting on the head-mounted display (HMD)). In MR, the sense of transportation can be similarly described when users select one dominant spatial (space-) reference frame. Then metaphors indicating a change of space would be similarly appropriate (e.g., portals). Thus, similar questions measuring spatial transportation might also be valid.
2) However, a reference framed by object-based *entities* draws attention to techniques that transport *others* to the user. Including the second type of transportation might refer to the feeling of sharing the room with objects. If this dimension is relevant for spatial presence in MR, measures should also draw attention to the other transportations. Questions asking for the sense of being might be supplemented by questions asking for the being with object-based *entities* or even space-related ones. Similarly, design considerations could think about transition techniques of object-based *entities* (e.g., fading in objects, replaying objects). Those techniques should correspond to the context to support a fluent experience. Vice versa, the affordance of the transportation technique might influence the set of the reference frame (if users need/have one). Thus, the affordance of transportation techniques might offer a vast design space for MR experiences. The time course of transportations might also be exciting in MR. Affordances of transportation at the beginning of the experience (when the user is forming expectations) might be different from transportations during the experiences (when users update expectations).
3) When users do not form a reference frame, the transportation dimension might be doubted in a general sense. Then, transportation (i.e., the *place-illusion*) might not be an appropriate metaphor for the user's experiences in MR, and consideration of new metaphors would probably be necessary.

RD3: Research Desiderata Concerning the Dimension of Realism in Mixed Realities

The dimension of *realism* refers to the *plausibility-illusion* (i.e., the sense of *being there plus*; Skarbez et al., 2017). In VR, plausibility addresses mainly objects, entities, and events that make sense in the place accepted as real (PERF). MR includes incongruences between the *space-based* and object-based *entities* by definition emphasizing the RD3, including the following research questions:

C-RQ 3.1: What can plausibility mean in MR defined by incongruences between the *space-based* and the object-based *entities*?
C-RQ 3.2: Which user expectancies or sets of (in-)coherence shape the *plausibility-illusion* in MR?
M-RQ 3.1: How can the *plausibility-illusion* (realism dimension) be measured in MR?
M-RQ 3.2: How can the presence of an (in-)coherent experience be measured in MR?
D-RQ 3.1: Which implications arise for the plausibility design in MR?
D-RQ 3.2: Which implications arise for an (in-)coherent design in MR?

The authors identify the following three views on the RD3.

1) The sense of plausibility probably refers to the views regarding the rest frame described above. In VR, the *plausibility-illusion* is closely related to the extent to which an *entity* makes sense in the place that has been accepted as real (PERF). In MR, the sense of plausibility can be similarly described when users select one dominant spatial (space-) reference frame. Then incongruent *entities* (i.e., *virtual objects* in a *real space* or *real objects* in a *virtual space*) would be perceived as plausible if they behave coherently to the dominant *place-illusion* (i.e., the selected *space-PERF*). For example, in AR, virtual objects should fall downwards and not hang in the air. Measures and design considerations might be similar to corresponding considerations in VR.
2) In contrast, when object-based *entities* determine the reference frame, plausibility should be evaluated relative to the object-based *entity*. Then *spaces* would be perceived as plausible if they behave coherently to the dominant *object-reference* (i.e., the selected PERF). Measures should probably assess coherence and plausibility relations between the object-based *entities*. Designers might draw attention to those relations when the goal is a fluent and coherent experience, even during transitions on the mRVC.
3) Suppose users do not need or have any reference frame. In that case, plausibility might be considered in a wider way or even in a different way. Particularly incoherent sets of circumstances might define the plausibility of MR. Measurements and design considerations should also reflect such a definition.

In sum, in MR, *object-based* entities might play an additional role in MR to the *space-based entities* in VR. Similar to in VR experiences, when users form a dominant reference frame, the question arises as to whether the *space-based* or the object-based *entities* determine it. Then, questions around the dimensions of transportation and realism could either be similar to VR or draw more attention to the object-based *entity*. Finally, users might not need or have a dominant reference frame or alternate it continuously. Then the construct of transportation could be doubted in general. The dimension of realism might also change since exceptionally incoherent sets of circumstances and not coherent sets might define the plausibility of MR.

The research desiderata and corresponding considerations lead up to the final question: Does spatial presence in MR refer to the sense of being anywhere (space-related, inside-out) or to the sense of being with something (object-related, outside-in) somewhere?

EXPERIMENTAL SETUPS TO TEST THE RESEARCH QUESTIONS

Hitherto, the discussion reflected spatial presence in MR. Diverse experimental setups could find answers to the research questions. Presenting one experiment for each question goes beyond the scope of the present article. Thus, we present one example to reveal the research heuristic behind such experiments. We hope that empirical studies convey the questions into a coherent set of assumptions, measurements, and useful design suggestions.

Control for Confounding Factors

Since immersion, defined as objective system factors, influences the *place-illusion* in VR (e.g., Skarbez et al., 2017; Wienrich and Gramlich, 2020), it might also play a significant role in MR (except the characteristic of inclusiveness). One good practice in VR research is to hold confounding immersive factors constant such as the kind of display. For example, when researchers want to investigate the sense of agency in dependency of the virtual embodiment, the same HMD should be worn in all embodiment conditions to control undesirable impacts of the HMD. Similarly, experiments investigating merging real and virtual entities in MR should follow the same good practice. Thus, if researchers explore the feeling of spatial presence in MR, the display properties of the system should remain the same over the application runtime to avoid confounding immersive factors. Currently, there is only one way to use the same system to represent both fully immersive virtual environments, as well as the real environments, and each possible merging in between. This possibility is the use of see-through HMDs, i.e., fully immersive VR glasses, which are able to display the real environment through cameras.

Paradigmatic Investigation of RD1–The Reference Frame in Mixed Realities

The questions about the reference frame in MR might be the most significant one of our considerations presented above. Hence, we present a paradigmatic experiment concerning the first research desideratum in the following.

The task of participants might be to search and merge objects by color. For that, they have to move around, find objects, and bring them to an object with the same color. During the experiment, participants would experience at least six transitions on the mRVC. Each transition reflects a systematic variation, either of a *space-based* or *object-based entity* (independent variable). To control for order-effects, the order of transitions would be balanced. One order of transitions is shown in **Table 1**. Before and after each transition, questions including the set, or the change of reference frame would be assessed (first-order dependent variable). Notably, the question must be assessed without switching off the HMD due to the above-described reasons. In addition, the sense of transportation (i.e., sense of *place-illusion*) or the sense of realism (i.e., sense of *plausibility-illusion*) could be assessed (second-order dependent variables).

The results of the experiments would bring primary answers to C-RQ 1.1 (*Do users have/need a dominant reference frame in MR experience?*) and C-RQ 1.2 (*Does the space-based or object-based entities determine the reference frame in MR experiences?*). Furthermore, the results would indicate how the sense of transportation (i.e., *place-illusion* in VR) or realism (i.e., *plausibility-illusion* in VR) are evaluated in MR.

DISCUSSION

Aim of the Present Considerations

Plenty of considerations, models, measures, and investigations target the understanding of the sense of presence in VR. However, full-immersive virtual reality is only one possible variant of the *mixed reality continuum* (Milgram and Kishino, 1994; Milgram and Colquhoun, 1999). Other variants have less immersive features (are less inclusive) and continuously interfere with the real physical space questioning the applicability of the accumulated knowledge about virtual presence and corresponding outcomes. The current work presents an alternative conceptional idea of reference setting in MR which raises the question: Does spatial presence in MR refer to the sense of being anywhere (space-related, inside-out) or to the sense of being with something (object-related, outside-in) somewhere? It also presents a research heuristic to investigate the idea and resulting question.

Contribution of the Present Considerations

To achieve the aim, we 1) presented a short review of definitions, dimensions, and measurements of presence in VR, and 2) presented the state of the art views on MR. Furthermore, we 3) derived a working definition of MR implying new conceptional ideas for the reference setting beyond full immersive VR and environmental cues. In our opinion, detaching from environmental entities as anchor cues is the essential assumption to discuss the spatial presence in MR. Based on these new ideas, we 4) identify three research desiderata, including research questions about the referential power of entities occurring in MR, the corresponding dimension of transportation (i.e., *place-illusion* in VR), and the dimension of

TABLE 1 | Shows one order of transitions of the paradigmatic experiment.

Transition	Space-based entities	*object-based entities*	Change of	Main questions
Basis 1	real	real	no	Does the participant ***set*** a (*space*-, *object* -) *PERF*?
Basis 1 A	real	virtual	*object-based entity*	Does the participant ***change*** the PERF?
Basis 1 B	virtual	real	space-based entity	Does the participant ***change*** the PERF?
Basis 1 C	virtual	virtual	space- and object-based *entity*	Does the participant ***change*** the PERF?
Basis 2	virtual	virtual		Does the participant ***set*** a (*space*-, *object*-) *PERF*?
Basis 2 A*	virtual	real	*object-based entity*	Does the participant ***change*** the PERF?
Basis 2 B*	real	virtual	space-based entity	Does the participant ***change*** the PERF?
Basis 2 C*	real	real	space- and object-based *entity*	Does the participant ***change*** the PERF?

realism (i.e., *plausibility-illusion* in VR) in MR. Finally, we 5) suggest an experimental setup to reveal the research heuristic behind experiments investigating presence in MR.

Working Definition of Mixed Realities

The working definition of MR is considered as concrete computer-generated Virtual Reality traversing the *modified Reality-Virtuality-Continuum* (short mRVC). Similar to the continuum defined by Milgram and Colquhoun (1999), the poles of the mRVC are defined as natural or *physical reality* (R), and *virtuality* (V). The positions between them refer to concrete forms of reality illusions (**Figure 1**). Each of these concrete forms of reality illusion defines a mixed reality illusion at one time point. Each reality illusion includes entities reaching from real to virtual manifestations at one time point. Entities can refer to spatial environments (i.e., space-based) or concrete objects (i.e., object-based). In addition, referential cues stemming from the presented entities themselves determine the selection of the reference frame. Our view overcomes the distinction of environments and objects by introducing entities possessing different degrees of referential power. It follows from the foregoing that spatial environmental entities and objects theoretically can serve as referential anchor cues. (**Figure 1**). Moreover, the mRVC uses a naive percent-visualization referring to the portion of real or virtual pictures (**Figure 2**). The amount classification of real and virtual picture portions might be supportive for systematic testing.

Research Desiderata and Research Questions About Spatial Presence in Mixed Realities

According to the inclusion of the *space-based* and *object-based entities* possessing referential power in the definition of MR, the question arises as to whether the spatial presence in MR refers to the sense of being anywhere (*space*) or to the sense of being with something (*object*) somewhere?

Three research desiderata underpin this question: RD1 concerns the reference frame, RD2 regards the dimension of transportation (i.e., *place-illusion* in VR), and RD3 refers to the realism dimension (i.e., *plausibility-illusion* in VR). Each desideratum includes three classes of research questions, C-RQ concerns the construct, M-RQ regards the measurements, and D-RQ refers to design challenges.

The authors detected three views on the RD1. Each further influenced the views on transportation and realism. The first view assumed the *place-illusion*, i.e., the self-positioning in a real or virtual environment, determines the PERF. Following this assumption, the user's positioning in a *real space* or *virtual space* determines the PERF selection. Then users would accept the *real* or *virtual objects* as a part of the space selected as *space-PERF*. The second view assumed that not the *space-based entities,* but the object-based *entities* determined the PERF. Then users would accept the *real* or *virtual spaces* as a part of the place selected as *object-PERF*. While both views assume one dominant reference frame, either *space* or the *object*, the third view assumed that users alter the PERF continuously or just do not have one. Then, space could be seen as the sum of its *objects*. In order to measure this, we would argue for including direct questions about the reference frame and the referential power of entities. One operationalization could be to assess the user's expectations or breaks in expectations relating to the *place-illusion* emerging by a real or virtual *space*, or the place spanned by the object-based *entities*. Those questions turn the balance when a real and a virtual PERF would make a huge difference. Results might be essential for design. When users form a dominant PERF, either by *space* or *objects*, MR applications should include easy to percept cues indicating clearly to a (*space*, *object-*) *R-PERF* or a (*space*, *object-*) *V-PERF*. In the case of no dominant PERF, many possibilities occur for the design that might be different from VR design.

Similarly, three views for the dimension of transportation (i.e., *place-illusion* in VR) and realism (i.e., *plausibility-illusion* in VR) are discussed. In MR, the sense of transportation can be similarly described to VR when users select one dominant spatial reference (*space-PERF*). Metaphors indicating a change of place would be similarly appropriate (e.g., portals). Similar questions measuring spatial transportation might also be valid. However, a reference framed by the object-based *entities* draws attention to techniques that transport others to the user. Measures should also draw attention to the other-transportations. Similarly, design considerations could involve transition techniques of the different *entities occurring in MR*. For both views, the affordance of the transportation technique, and vice versa, might influence the set of the reference frame. Thus, the affordance of transportation techniques might offer a vast design space for MR experiences. In addition, the time course of transportations might also be an exciting topic for MR design. In contrast, the third view doubted the existence of a transportation dimension in the case users do not form a dominant reference frame. Thus, transportation (i.e., the *place-*

illusion in VR) might not be an appropriate metaphor for the user's experiences in MR, and consideration of new metaphors would probably be necessary.

In MR, the sense of plausibility can be similarly described to VR when users select one dominant space-based reference frame. Then even incongruent *entities* (i.e., *virtual objects* in a *real spaces* or *real objects* in *virtual spaces*) could be perceived as plausible if they behave coherently with the dominant *space-PERF*. In contrast, when *object-based entities* determine the reference frame, plausibility should be evaluated relative to them. Then *spaces* would be perceived as plausible if they behave coherently to the dominant *object-PERF*. We argued for measures and designs similar to VR in the former case and supplementing the assessment and design of coherence and plausibility relations between the *entities* in the latter. Suppose users do not need or have any frame of reference. In that case, the dimension of realism (i.e., *plausibility-illusion* in VR) might also change since exceptionally incoherent sets of circumstances and incoherent sets might define the plausibility of MR.

In sum, the present work contributes to the debate about what learnings and knowledge about presence and corresponding outcomes collected in the context of VR can be gainful for research in MR. The research questions considered guiding questions for experimental setups resulting in substantiation of spatial presence in MR. One paradigmatic experiment is described to illustrate a possible research heuristic for future work.

Limitations and Future Work

Five main limitations characterize the present contribution.

1) The first limitation refers to the focus on spatial presence. As mentioned above, presence is a broad construct, including, besides spatial presence (focus here), social presence (e.g., Lee, 2004), or cognitive presence (Nunez and Blake, 2001), for example. Furthermore, this paper considers spatial presence in the context of technologically mediated realities such as VR, AR, AV, and MR and excludes senses caused by dreams or drugs. Consequently, the discussions are restricted to the conceptional focus set by current debates of spatial presence in VR. The *place-illusion* and *plausibility-illusion* (Slater, 2009; Skarbez et al., 2017) and corresponding dimensions (Lombard and Ditton, 1997) combined with the idea of a *primary (space-based) egocentric reference frame* (Wirth et al., 2007) guided the presented research questions. Thus, scrutinizing current views on presence in VR was out of the scope here. Future work should examine other sub-constructs that refer to presence to consider their application for evaluating MR experiences. Similarly, revisited views on the concept of presence, such as the view of Latoschik and Wienrich (2021) should be incorporated in the future.
2) Second, the present considerations are limited to the working definition of MR. The working definition is based on the current scientific view on MR (e.g., Milgram and Kishino, 1994; Milgram and Colquhoun, 1999; Speicher et al., 2019). Other definitions, such as industrial views (Microsoft, 2020), were not in the scope of the present paper. In addition, Milgram and Colquhoun (1999) extended their definition of MR in the form of transitions between reality and virtuality (i.e., mixed reality combination spaces). This extension opens up the MR view, particularly for those allowing for transitions, such as MR headsets' see-through function. A more radical interpretation of the transitions might be the inclusion of transition, i.e., interaction between the respective reality illusions on the continuum, as a necessary part of the reality form, which is called MR. Thus, MR is not an umbrella term but a specific type of reality similar to and consisting of AR and AV. MR's distinct feature would be an interpolation between the extreme *reality* and *virtuality* on the continuum. An MR application is one in which within the runtime of one application and within one single display, the user traverses from reality to virtuality or vice versa at least once by stepping through instances of R, AR, AV, and VR frames. If you stopped the application at any point in time, the frame could be described as either R, AR, AV, or VR. Nevertheless, researchers need a concrete definition of MR and future work should examine whether the present considerations are valid for different MR definitions or dimensions, as introduced recently by Speicher and colleagues (2019), for example.
3) The present considerations include single-user MR applications. However, in the future, MR applications will probably allow for multi-user functions. Questions about joint reference frames and anchors will then arise. Moreover, MR applications allowing personalized viewing content, such as glasses showing personalized advertisements in the supermarket, cause questions about different reference frames for different users within the same application. Future work should examine whether the present considerations are valid and incorporate different views for those applications.
4) Although the present work considers implications for measuring and designing MR experience, the focus is laid on the conceptional implications. In addition, the research heuristic is only described by one paradigmatic experiment. Thus, empirical and practical validity is limited. In our opinion, implications for measures and design result from conceptional considerations, and practical research needs a starting point. However, future studies should focus on empirical data and practical significance.
5) Finally, the present work assumed that MR experiences aim for fluent experiences. Thus, plausibility can be linked to coherence (Skarbez et al., 2017). The zeitgeist focuses on intuitive use and technical usage without effort (Hartson and Pyla, 2012). However, intuitive use can also lead to uncritical usages and the psychological risks of misunderstanding technical devices and their power (Long and Magerko, 2020). The design of safety-critical systems already includes so-called intentional frictions or desirable difficulties (Druckman and Bjork, 1994). Thus, when using MR in safety-critical contexts (e.g., hospitals) disfluent experiences might also have relevant applications. Future work should examine plausibility and the link to coherence for those applications.

CONCLUSION

The present article analyzed the construct of spatial presence in MR. It presented an alternative conceptional idea of reference setting in MR, which raises the question: Does spatial presence in MR refer to the sense of being anywhere (space-related, inside-out) or to the sense of being with something (object-related, outside-in) somewhere? The current work also presented a research heuristic to investigate the idea and resulting question. Considerations about implications for the concept, the measurement, and the design of spatial presence in MR are encouraged. We hope further that empirical studies, described paradigmatically, convey the questions into a coherent set of assumptions, measurements, and valuable design suggestions. The construct of virtual presence is strongly linked with user experience in Virtual Reality (e.g., Skarbez et al., 2018). The feeling of being there (i.e., virtual spatial presence) can probably be considered as so-called *hygiene factors* (Wienrich and Gramlich, 2020). To a certain extent, it might be necessary to allow other VR potentials to become effective. It might be conceptualized similarly to the role of pragmatic quality within the field of user experience (Hassenzahl et al., 2010). For MR experiences, the research about essential underlying factors, hygiene factors, determining the user experience is still in its infancy. The present considerations might be a promising starting point.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

CW was the main idea giver, made the theoretical concept, made the empirical concept, was the main writer PK was an idea giver, contributed to the theoretical and empirical concept SV supported in writing, contributed to the empirical concept ML contributed mainly to the idea and to the theoretical concept and supported writing a lot.

FUNDING

This publication was supported by the Open-Access Publication Fund of the University of Wuerzburg.

REFERENCES

Alexander, A. L., Brunyé, T., Sidman, J., and Weil, S. A. (2005). From Gaming to Training: A Review of Studies on Fidelity, Immersion, Presence, and Buy-In and Their Effects on Transfer in Pc-Based Simulations and Games. *DARWARS Train. Impact Group* 5, 1–14.

Billinghurst, M., Kato, H., and Poupyrev, I. (2001). The Magicbook-Moving Seamlessly between Reality and Virtuality. *IEEE Comp. Graphics Appl.* 21 (3), 6–8. doi:10.1109/38.920621

Biocca, F. (2003). "Can We Resolve the Book, the Physical Reality, and the Dream State Problems? from the Two-Pole to a Three-Pole Model of Shifts in Presence," in EU Future and Emerging Technologies, Presence Initiative Meeting (Venice, Italy: Citeseer), 13–69. doi:10.1162/105474603322761270

Brown, S., Ladeira, I., Winterbottom, C., and Blake, E. (2003). "The Effects of Mediation in a Storytelling Virtual Environment," in International Conference on Virtual Storytelling (Springer), 102–111. doi:10.1007/978-3-540-40014-1_13

Druckman, D. E., and Bjork, R. A. (1994). *Learning, Remembering, Believing: Enhancing Human Performance*. Washington, DC: National Academy Press.

Gibson, J. J. (1966). *The Senses Considered as Perceptual Systems*. Washington: American Psychological Association.

Hartson, R., and Pyla, P. S. (2012). *The UX Book: Process and Guidelines for Ensuring a Quality User Experience*. Elsevier.

Hassenzahl, M., Diefenbach, S., and Göritz, A. (2010). Needs, Affect, and Interactive Products - Facets of User Experience. *Interacting Comput.* 22 (5), 353–362. doi:10.1016/j.intcom.2010.04.002

Hillstead, A. (2017). *Simulating Psychedelic Therapy through Mediated Reality*. Wellington: Victoria University of Wellington.

Insko, B. E. (2003). *Measuring Presence: Subjective, Behavioral and Physiological Methods*. Washington: American Psychological Association.

Latoschik, M. E., and Wienrich, C. (2021). *Coherence and Plausibility, Not Presence?! Pivotal Conditions for XR Experiences and Effects, a Novel Model*. New York: ArXiv Preprint.

Lee, K. M. (2004). Presence, Explicated. *Commun. Theor.* 14 (1), 27–50. doi:10.1111/j.1468-2885.2004.tb00302.x

Lee, S., Kim, G. J., Rizzo, A., and Park, H. (2004). "Formation of Spatial Presence: By Form or Content?," in Proceedings of the 7th International Workshop on Presence, 20–27.

Lessiter, J., Freeman, J., Keogh, E., and Davidoff, J. (2001). A Cross-media Presence Questionnaire: The ITC-Sense of Presence Inventory. *Presence: Teleoperators & Virtual Environments* 10 (3), 282–297. doi:10.1162/105474601300343612

Lombard, M., and Ditton, T. (1997). At the Heart of it All: The Concept of Presence. *J. Computer-Mediated Commun.* 3 (2), JCMC321. doi:10.1111/j.1083-6101.1997.tb00072.x

Long, D., and Magerko, B. (2020). "What Is AI Literacy? Competencies and Design Considerations," in Conference on Human Factors in Computing Systems - Proceedings (Honolulu: Association for Computing Machinery), 1–16. doi:10.1145/3313831.3376727

Maguire, K. C., and Connaughton, S. L. (2006). "Relational Presence and Distanced Interdependent Relationships," in *The Ninth International Workshop on Presence* (Cleveland, OH: PRESENCE2006).

Mann, S. (2002). *Mediated Reality with Implementations for Everyday Life*. Cambridge, MA: MIT Press journal PRESENCE, 1.

Mann, S., and Nnlf, S. M. (1994). *Mediated Reality*. Philadelphia, PA: cyborganthropology.

Microsoft (2020). *Was its Mixed Reality?* Philadelphia: The Franklin Institute.

Milgram, P., and Colquhoun, H. (1999). A Taxonomy of Real and Virtual World Display Integration. *Mixed Reality: Merging Real and Virtual Worlds* 1, 1–26. doi:10.1007/978-3-642-87512-0_1

Milgram, P., and Kishino, F. (1994). A Taxonomy of Mixed Reality Visual Displays. *IEICE Trans. Inf. Syst.* E77-D (12), 1. 10.1.1.102.4646

Minsky, M. (1980). Telepresence. *Omni* 2 (9), 1.

Nunez, D. (2007). *A Capacity Limited, Cognitive Constructionist Model of Virtual Presence*. Town: University of Cape.

Nunez, D., and Blake, E. (2001). "Cognitive Presence as a Unified Concept of Virtual Reality Effectiveness," in Proceedings of the 1st International Conference on Computer Graphics, Virtual Reality and Visualisation, 115–118. doi:10.1145/513867.513892

Schubert, T., Friedmann, F., and Regenbrecht, H. (2001). The Experience of Presence: Factor Analytic Insights. *Presence: Teleoperators & Virtual Environments* 10 (3), 266–281. doi:10.1162/105474601300343603

Schubert, T., Regenbrecht, H., and Friedmann, F. (2016). *Igroup Presence Questionnaire*. IPQ.

Schubert, T. W. (2009). A New conception of Spatial Presence: Once Again, with Feeling. *Commun. Theor.* 19 (2), 161–187. doi:10.1111/j.1468-2885.2009.01340.x

Schubert, T. W. (2003). The Sense of Presence in Virtual Environments:. *Z. für Medienpsychologie* 15 (2), 69–71. doi:10.1026//1617-6383.15.2.69

Schwind, V., Knierim, P., Haas, N., and Henze, N. (2019). "Using Presence Questionnaires in Virtual Reality," in Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 1–12. doi:10.1145/3290605.3300590

Sheridan, T. B. (1992). Musings on Telepresence and Virtual Presence. *Presence: Teleoperators & Virtual Environments* 1 (1), 120–126. doi:10.1162/pres.1992.1.1.120

Skarbez, R., Brooks, F. P., Jr, and Whitton, M. C. (2018). "Immersion and Coherence in a Stressful Virtual Environment," in Proceedings of the 24th Acm Symposium on Virtual Reality Software and Technology, 1–11. doi:10.1145/3281505.3281530

Skarbez, R., Brooks, Jr., F. P., and Whitton, M. C. (2018). A Survey of Presence and Related Concepts. *ACM Comput. Surv.* 50, 1–39. doi:10.1145/3134301

Slater, M. (1999). "Co-presence as an Amplifier of Emotion," in 2nd International Workshop on Presence (Colcester, England: University of Essex).

Slater, M., McCarthy, J., and Maringelli, F. (1998). The Influence of Body Movement on Subjective Presence in Virtual Environments. *Hum. Factors* 40 (3), 469–477. doi:10.1518/001872098779591368

Slater, M. (2009). Place Illusion and Plausibility Can lead to Realistic Behaviour in Immersive Virtual Environments. *Phil. Trans. R. Soc. B* 364 (1535), 3549–3557. doi:10.1098/rstb.2009.0138

Slater, M., Sadagic, A., Usoh, M., and Schroeder, R. (2000). Small-group Behavior in a Virtual and Real Environment: A Comparative Study. *Presence: Teleoperators & Virtual Environments* 9 (1), 37–51. doi:10.1162/105474600566600

Slater, M., and Steed, A. (2000). A Virtual Presence Counter. *Presence: Teleoperators & Virtual Environments* 9 (5), 413–434. doi:10.1162/105474600566925

Slater, M., and Usoh, M. (1993). Simulating Peripheral Vision in Immersive Virtual Environments. *Comput. Graphics* 17 (6), 643–653. doi:10.1016/0097-8493(93)90113-n

Slater, M., and Wilbur, S. (1997). A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. *Presence: Teleoperators & Virtual Environments* 6 (6), 603–616. doi:10.1162/pres.1997.6.6.603

Speicher, M., Hall, B. D., and Nebeling, M. (2019). "What Is Mixed Reality?," in Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 1–15.

Steuer, J. (1992). Defining Virtual Reality: Dimensions Determining Telepresence. *J. Commun.* 42 (4), 73–93. doi:10.1111/j.1460-2466.1992.tb00812.x

Usoh, M., Arthur, K., Whitton, M. C., Bastos, R., Steed, A., Slater, M., et al. (1999). "Walking> Walking-In-Place> Flying, in Virtual Environments," in Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques, 359–364. doi:10.1145/311535.311589

Usoh, M., Catena, E., Arman, S., and Slater, M. (2000). Using Presence Questionnaires in Reality. *Presence: Teleoperators & Virtual Environments* 9 (5), 497–503. doi:10.1162/105474600566989

Van Schaik, P., Turnbull, T., Van Wersch, A., and Drummond, S. (2004). Presence within a Mixed Reality Environment. *Cyberpsychol Behav.* 7 (5), 540–552. doi:10.1089/1094931042403145

Varjo (2019). *Unmodified XR-1 Footage*. Helsinki, Finland: Vimeo.

Wagner, I., Broll, W., Jacucci, G., Kuutii, K., McCall, R., Morrison, A., et al. (2009). On the Role of Presence in Mixed Reality. *Presence: Teleoperators and Virtual Environments* 18 (4), 249–276. doi:10.1162/pres.18.4.249

Wienrich, C., and Gramlich, J. (2020). appRaiseVR - an Evaluation Framework for Immersive Experiences. *I-Com* 19 (2), 103–121. doi:10.1515/icom-2020-0008

Wirth, W., Hartmann, T., Böcking, S., Vorderer, P., Klimmt, C., Schramm, H., et al. (2007). A Process Model of the Formation of Spatial Presence Experiences. *Media Psychol.* 9 (3), 493–525. doi:10.1080/15213260701283079

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

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Virtual Reality

ORIGINAL RESEARCH
published: 25 October 2021
doi: 10.3389/frvir.2021.695312

Move The Object or Move The User: The Role of Interaction Techniques on Embodied Learning in VR

Mahda M. Bagher[1], Pejman Sajjadi[1]*, Jan Oliver Wallgrün[1], Peter C. La Femina[2] and Alexander Klippel[1]*

[1]Center for Immersive Experiences, Department of Geography, The Pennsylvania State University (PSU), University Park, PA, United States, [2]Department of Geosciences, The Pennsylvania State University, University Park, PA, United States

To incorporate immersive technologies as part of the educational curriculum, this article is an endeavor to investigate the role of two affordances that are crucial in designing embodied interactive virtual learning environments (VLEs) to enhance students' learning experience and performance: 1) the sense of presence as a subjective affordance of the VR system, and 2) bodily engagement as an embodied affordance and the associated sense of agency that is created through interaction techniques with three-dimensional learning objects. To investigate the impact of different design choices for interaction, and how they would affect the associated sense of agency, learning experience and performance, we designed two VLEs in the context of penetrative thinking in a critical 3D task in geosciences education: understanding the cross-sections of earthquakes' depth and geometry in subduction zones around the world. Both VLEs were web-based desktop VR applications containing 3D data that participants ran remotely on their own computers using a normal screen. In the drag and scroll condition, we facilitated bodily engagement with the 3D data through object manipulation, object manipulation. In the first-person condition, we provided the ability for the user to move in space. In other words, we compared moving the objects or moving the user in space as the interaction modalities. We found that students had a better learning experience in the drag and scroll condition, but we could not find a significant difference in the sense of presence between the two conditions. Regarding learning performance, we found a positive correlation between the sense of agency and knowledge gain in both conditions. In terms of students with low prior knowledge of the field, exposure to the VR experience in both conditions significantly improved their knowledge gain. In the matter of individual differences, we investigated the knowledge gain of students with a low penetrative thinking ability. We found that they benefited from the type of bodily engagement in the first-person condition and had a significantly higher knowledge gain than the other condition. Our results encourage in-depth studies of embodied learning in VR to design more effective embodied virtual learning environments.

Keywords: virtual reality, embodied learning, embodiment, bodily engagement, interaction technique, virtual learning environments, penetrative thinking, 3D visualization

OPEN ACCESS

Edited by:
Amela Sadagic,
Naval Postgraduate School,
United States

Reviewed by:
Andri Ioannou,
Cyprus University of Technology,
Cyprus
Missie Smith,
Independent researcher, Detroit, MI,
United States

****Correspondence:***
Mahda M. Bagher
mmm6749@psu.edu
Pejman Sajjadi
sfs5919@psu.edu

Specialty section:
This article was submitted to Virtual Reality and Human Behaviour, a section of the journal Frontiers in Virtual Reality

Received: *14 April 2021*
Accepted: *22 September 2021*
Published: *25 October 2021*

Citation:
Bagher MM, Sajjadi P, Wallgrün JO, La Femina PC and Klippel A (2021) Move The Object or Move The User: The Role of Interaction Techniques on Embodied Learning in VR. Front. Virtual Real. 2:695312. doi: 10.3389/frvir.2021.695312

1 INTRODUCTION

Extended Reality (XR) technologies have become more accessible in terms of costs and required hardware and software and have gained attention and popularity in education (e.g., Dalgarno et al., 2011; Bulu, 2012; Merchant et al., 2014; Legault et al., 2019; Klippel et al., 2019). Recent advances in XR technologies have created an interest in investigating the role of cognitively motivated principles in designing virtual learning environments (VLEs) for education (e.g., Dalgarno and Lee, 2010; Lee et al., 2010; Johnson-Glenberg et al., 2014; Clifton et al., 2016; Yeonhee, 2018). There have been numerous efforts from various communities (e.g., IEEE ICICLE[1] and The Immersive Learning Research Network (iLRN)[2]) to incorporate the technology-enhanced educational curriculum into classrooms, to overcome the limitations of learning technologies, and to design engaging and compelling learning experiences. The learning efficacy of these experiences is a product of their design, which in turn predicts the experiences of users (Dalgarno and Lee, 2010; Clifton et al., 2016; Jerald, 2016; Czerwinski et al., 2020). Among the various aspects that should be considered when designing an interactive virtual environment for learning, embodiment is argued to be one of the main contributors (Biocca, 1999; Johnson-Glenberg, 2018; Johnson-Glenberg et al., 2020). Within a rich body of research on the role of embodiment in spatial learning, thinking, and reasoning (e.g., Mou and McNamara, 2002; Wilson, 2002; Hegarty et al., 2006; Hostetter and Alibali, 2008; Kelly and McNamara, 2008; Kelly and McNamara, 2010; Paas and Sweller, 2012; Shapiro, 2014; Plummer et al. (2016)), there is a growing interest in investigating the role of embodiment in the design of VLEs as an essential factor influencing immersive learning (e.g., Kilteni et al., 2012; Lindgren and Johnson-Glenberg, 2013; Johnson-Glenberg et al., 2014; Lindgren et al., 2016; Clifton et al., 2016; Johnson-Glenberg, 2018; Skulmowski and Rey, 2018; Legault et al., 2019; Johnson-Glenberg et al., 2020; Southgate, 2020; Bagher, 2020).

This growing body of research examines the extent to which embodied learning in a virtual environment would enhance learning outcomes and improve learners' spatial memory. Researchers in various fields have defined embodiment in different ways (Kilteni et al., 2012) and focused on numerous aspects, from body representation to the type of bodily engagement or the degree of embodiment. One common goal is to find out what type or degree of embodiment is beneficial in designing engaging and effective learning experiences in XR, especially virtual reality (Kilteni et al., 2012; Repetto et al., 2016; Johnson-Glenberg et al., 2016; Skulmowski and Rey, 2018; Johnson-Glenberg, 2018; Southgate, 2020; Johnson-Glenberg et al., 2020).

In this article, our focus is not the degree of embodiment but one of the *affordances* that play a key role in inducing the sense of embodiment (SOE) in VR. We investigate the extent to which bodily engagement (as an embodied affordance) contributes to SOE in VLEs and can affect learning experience and performance. Affordances are defined as "potential interactions with the environment" (Wilson, 2002, p.625). Different VR systems can afford different levels of *sensorimotor contingencies* depending on the system characteristics and the design choices for creating the learning environment. Sensorimotor contingencies refer to when we take certain actions to change our perception and interact with an environment, including but not limited to a virtual environment (Lee, 2004; Slater, 2009; Slater et al., 2010; Skulmowski and Rey, 2018). Johnson-Glenberg et al. (2014) refer to this as *motor engagement*. In this article, we use the term *bodily engagement* suggested by Skulmowski and Rey (2018) as this term entails a type of engagement that extends beyond the mind and considers the interaction between mind, body, and the environment (Wilson, 2002; Skulmowski and Rey, 2018). When the learning activities in a virtual environment are designed to engage the senses (i.e., vision) and motor engagement (i.e., body parts), the users experience higher engagement with those activities. As a result, they can be more embodied in the environment (Biocca, 1999; Jerald, 2016). The level of bodily engagement depends on the number of sensory systems engaged and whether the tasks are designed around meaningful activities. Bodily engagement can further affect memory trace and knowledge gain (Johnson-Glenberg et al., 2016; Skulmowski and Rey, 2018).

To examine the effect of bodily engagement on learning experience and performance, we focus on the design choices for bodily engagement in the same learning context with the same level of embodiment rather than evaluating the medium effect on learning. We have designed an experiment with two VLEs. These VLEs are web-based desktop VR applications. Web-based desktop VR refers to a desktop VR experience perceived via a standard screen delivered via a web browser. We argue that the type of 3D interaction for manipulation of virtual objects matters (Weise et al., 2019). In a recent study by Johnson-Glenberg et al. (2021) comparing immersive VR and a desktop VR with two levels of embodiment (low: passive video watching, high: interacting with the learning content), they found that the design is far important than the platform. The critical finding is that the way a learning environment is designed based on the presence or absence of interaction techniques matters in learning.

To carry out this research, first, we investigate the following questions: Does the type of interaction technique affect the level of bodily engagement and associated sense of agency? And does the type of interaction technique affect the sense of presence? To answer these questions, we look into 1) bodily engagement through two different interaction techniques and the associated sense of agency, and 2) the created sense of presence as the subjective or psychological affordance of the VR system (Slater and Wilbur, 1997; Ruscella and Obeid, 2021). We hypothesize that the design choices for the interaction technique influence the level of bodily engagement and the level of control over the learning environment that creates the sense of agency. This sense of agency can further affect the overall experienced sense of presence (Nowak and Biocca, 2003). Furthermore, presence, in return, has an effect on the level of bodily engagement and learning in VR (Johnson-Glenberg,

[1]https://sagroups.ieee.org/icicle/

[2]https://immersivelrn.org/about-us/what-is-ilrn/

2018). Extensive research has been carried out on the sense of presence as a psychological affordance of a VR system (e.g., Slater and Wilbur, 1997; Witmer and Singer, 1998; Schuemie et al., 2001; Lee, 2004; Sanchez-Vives and Slater, 2005; Wirth et al., 2007; Schubert, 2009; Slater et al., 2010; Bulu, 2012; Bailey et al., 2012)

The goal of the VLEs used in this study is to support penetrative thinking in the "Discovering Plate Boundaries[3]" lab in an introductory physical geology course. In short, penetrative thinking is the ability to visualize a 2D profile of three-dimensional data. In designing and incorporating the VLEs into the plate boundaries lab exercise, we explore these research questions: Do interaction techniques affect learning experience and performance? And is one interaction technique superior to the other for students with a low penetrative thinking ability in terms of knowledge gain? We hypothesize that the interaction technique affects the learning experience and performance in the context of penetrative thinking in VR as a type of spatial learning. In a pilot study (Bagher et al., 2020) conducted in the Fall 2019, we focused on the 3D visualization of the US Geological Survey's Centennial Earthquake Catalog (Ritzwoller et al., 2002) as a case study and immersive VR (IVR) using Head-Mounted Displays (HMDs) as an embodied and interactive learning experience. The pilot study focused on comparing IVR with the traditional teaching approach (using 2D maps) to determine whether IVR as an interactive 3D learning environment is superior to the traditional teaching methods. Due to the unprecedented event of the epidemic of COVID-19 during Fall 2020, physical attendance at the labs and using VR headsets (HMDs) was affected. Therefore, we created two VLEs based on virtual web-based desktop applications that presented the 3D visualization of the earthquake locations on a 2D interface with different interaction techniques. The use of a web browser was to give accessibility to students to attend the experiment from home. We incorporated the virtual learning environments into the curriculum to teach plate boundaries and earthquake locations, and they were the only method of learning available for the lab exercise. Therefore, this study explores whether the design of the interaction techniques used in the VLEs would affect learning experience and performance when VR is the established method of learning in the lab.

In the rest of the article, we first discuss the background of our research. Then, we discuss the design and implementation of the experiment. After reporting the results, we discuss their implications on learning experience, user experience, and learning performance. Then we address the limitations of the study and future directions for this research.

2 BACKGROUND

2.1 Sense of Embodiment

Embodied learning theory (Stolz, 2015; Smyrnaiou et al., 2016), as a pedagogical approach rooted in embodied cognitive science, seeks to expand the application of embodied cognition into education. Embodiment is experiencing and interacting with the world through our bodies, suggesting that mind and body are linked (Wilson, 2002; Kilteni et al., 2012; Smyrnaiou et al., 2016). Therefore, in contrast to traditional cognitive science, embodied cognition explains how body and environment are related to cognitive processes (Barsalou, 1999; Barsalou, 2008; Shapiro, 2007; Shapiro, 2014; Skulmowski and Rey, 2018). Embodiment is rooted in human perception and motor systems and through the body's interaction with the world rather than only relying on abstract symbolic and internal representations (Barsalou, 1999; Wilson, 2002; Waller and Greenauer, 2007; Shapiro, 2007, Shapiro, 2014). In recent years, the design of embodied interfaces, including immersive experiences, has captured the attention of researchers in different fields in an attempt to improve embodied learning (e.g., Dalgarno and Lee, 2010; Johnson-Glenberg et al., 2014; Clifton et al., 2016; Yeonhee, 2018; Czerwinski et al., 2020). To conceptualize embodiment in the context of virtual reality, we should define how SOE is constructed based on embodied mental representations. SOE is a psychological response to being situated in the space in relation to other objects and the self. A virtual interface can be an extension of human senses linking the human to the virtual environment (Biocca, 1999; Kilteni et al., 2012). In other words, SOE in VR can be defined as the integration of our senses with our technology extended bodies (Biocca, 1999).

Among research studies focused on embodiment in VR, some have focused on defining different contributing factors to the embodiment. For instance, Kilteni et al. (2012) define the sense of embodiment as a result of the sense of self-location, the sense of agency, and the sense of body ownership. Some researchers (e.g., Gonzalez-Franco and Peck, 2018) focus on the role of the body as an avatar and its effect on the sense of body ownership and agency. In another example, Southgate (2020) conceptualizes embodiment in virtual learning from different angles focusing on various representations of the body such as cyborg body, naturalistic body, political body, etc. Furthermore, several research studies are focusing on the role of bodily engagement on SOE in VR (e.g., Johnson-Glenberg, 2018; Skulmowski and Rey, 2018; Johnson-Glenberg et al., 2020; Johnson-Glenberget al., 2021). Johnson-Glenberg et al. (2020) defined two affordances for designing VR for learning: 1) the sensation of presence, and 2) embodiment and the agency linked with manipulating objects in 3D. They define embodiment as a meaningful interaction with the learning content through bodily engagement. In another study by Johnson-Glenberg et al. (2016), they found that embodiment and sensorimotor feedback can increase knowledge retention in some types of knowledge. Johnson-Glenberg et al. (2021) compared passive learning (watching a video) vs. active learning through embodied interactions on a 2D platform and an immersive VR (Oculus Go). In all conditions, users sit. In the active learning scenario, using a mouse on a 2D desktop and controllers in an immersive VR platform is highly embodied. Watching a video on both platforms is considered low embodied. Therefore, the user has the same level of bodily engagement both in VR and a 2D desktop when assigned to active learning. They found a

[3]Plate boundaries are the edges of plates created when the lithosphere is broken into multiple pieces (Tarbuck et al., 1997).

significant main effect for embodiment regardless of the platform. Participants in high embodied conditions learned the most. Zielasko and Riecke (2021) carry out a systematic analysis with VR experts in a workshop to find out the effect of body posture and embodied interactions on various VR experiences such as engagement, enjoyment, comfort, and accessibility. They also found higher embodied locomotion cues for walking rather than sitting. Among other research studies focusing on interaction techniques, locomotion, and embodiment (e.g., Zielasko et al., 2016; Weise et al., 2019; Di Luca et al., 2021), Lages and Bowman (2018) focused on the effect of manipulating objects vs physically walking in the virtual environment on performance in demanding visual tasks. They found that in designing the learning environments, the creator should consider the user controller experience, past gaming experience, and spatial ability of the user.

In a desktop VR, hands movement and a mouse or a keyboard simulate bodily engagement at a lower level, giving the user the sense of being situated in the virtual environment while sitting in front of a 2D interface. We consider this form of SOE as the lower level of bodily engagement than immersive VR, where the whole body can be moved and engaged. In this article, instead of comparing the degree of embodiment, we investigate the design choices for bodily engagement in two web-based desktop VR with the same level of embodiment. We posit that different design choices for interaction techniques would affect learning experience and performance. We hypothesize that various interaction techniques can generate different levels of agency over the learning materials and result in different learning outcomes in terms of knowledge gain. Two main interaction techniques with the learning contents introduced in the literature are 1) *gesture*, and 2) *object manipulation* (Paas and Sweller, 2012). Several studies have explored the role of gesture as an effective bodily engagement technique in learning spatial information and offloading mental tasks to the surrounding environment (e.g., Hostetter and Alibali, 2008; Lindgren and Johnson-Glenberg, 2013; Plummer et al., 2016; Johnson-Glenberg, 2018). We propose to add a third interaction technique, 3) *to move the user in space*. This interaction technique creates a sense of embodied locomotion and gives the user the ability to control the rotation of the viewpoint by either stepping back in x,y,z direction and seeing an overview of the 3D objects or moving closer to inspect the 3D objects in greater detail. We are interested in examining the role of object manipulation and moving the user in space as interaction techniques contributing to bodily engagement in enhancing learning and the associated sense of agency.

Bodily Engagement Through Object Manipulation

This interaction technique creates a sense of agency and control over the 3D objects in a three-dimensional environment. According to Paas and Sweller (2012) object manipulation is a source of primary knowledge that will not affect cognitive load during the learning process. The primary systems can further assist the user in acquiring secondary knowledge. Manipulating an environment can help us to solve a problem through mental structures that assists perception and action. Moreover, adding a modality like object manipulation in the immediate environment may increase the strength of memory trace and recall (Barsalou, 1999; Wilson, 2002; Johnson-Glenberg et al., 2016; Johnson-Glenberg et al., 2020). In the recall process, in the absence of physical activity, the sensorimotor actions like object manipulation can later assist the processes of thinking and knowing by representing information or drawing inferences (Barsalou, 1999; Wilson, 2002). Working memory has a sensorimotor nature and benefits from off-loading information into perceptual and motor systems (Wilson, 2002). Therefore, we suggest using object manipulation to help with the cognitive load that can increase working memory capacity. Object manipulation in a web-based desktop VR can be achieved through dragging, rotation, and scroll using a mouse. Many 3D software programs use this technique to manipulate 3D content.

Bodily Engagement Through Moving the User in Space

Moving in space either physically in a virtual environment or through controller-based navigation in a web-based desktop VR is a cognitively demanding task. Changing perspective to create a different perception of the environment to perform a task or solve a problem is called epistemic action (Hostetter and Alibali, 2008). Epistemic actions are the result of sensorimotor contingencies (Slater, 2009; Slater et al., 2010) supported by a VR system. Even though physically walking is considered to be cognitively demanding, it is considered to be the most natural interaction technique (Lages and Bowman, 2018). Zielasko and Riecke (2021) carry out a survey in which participants rated higher embodied (non-visual) locomotion cues for walking, walking in place, and arm swinging than standing, sitting, or teleportation. In a web-based desktop VR, physical walking can be replicated using a controller. Moving in space can benefit from familiarity with controller-based games (Lages and Bowman, 2018) such as First Person Shooter (FPS) games. In these games, the player has an egocentric view and controls the movement in space in different directions using a game controller device or a mouse and keyboard.

2.2 Penetrative Thinking

Spatial thinking is a fundamental part of many fields of science. One of the ways students can gain a better understanding of a spatial phenomenon is through visual-spatial thinking (Mathewson, 1999). Adequate visualization helps students to understand the spatial representation of information better. Spatial representations can be either extrinsic (e.g., locations) or intrinsic (e.g., shapes) to objects. One of the important spatial transformations related to intrinsic characteristics of objects is the ability to visualize penetrative views and to switch between two-dimensional and three-dimensional views. The ability to understand spatial relations inside an object and transform 3D data into a 2D profile is called *penetrative thinking* or *cross-sectioning* (Ormand et al., 2014; Newcombe and Shipley, 2015; Hannula, 2019). **Figure 1** shows a penetrative thinking ability test to test students' ability on mental slicing of a 3D geologic structure in a block diagram (Ormand et al., 2014).

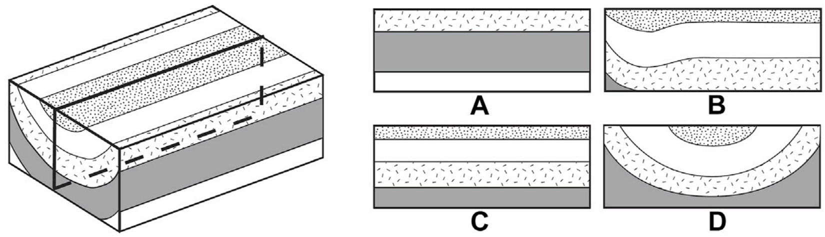

FIGURE 1 | Geologic Block Cross-sectioning Test for measuring students' ability on mental slicing of a 3D geologic structure in a block diagram. The GBCT post-study test re-published from (Ormand et al., 2014).

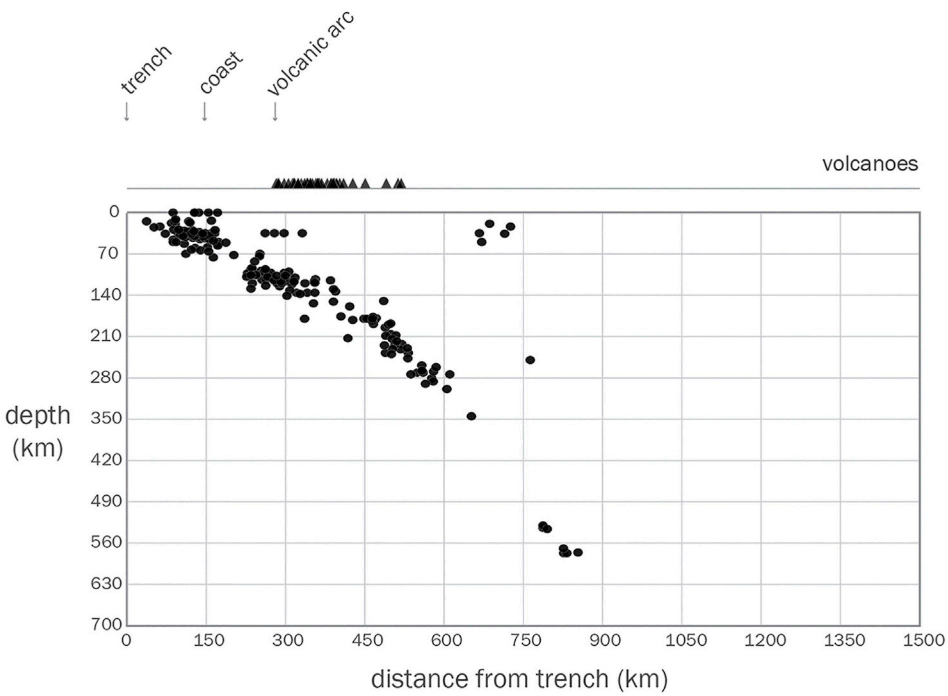

FIGURE 2 | An example of a plot drawn in an introductory geoscience course: cross-section of earthquakes and volcanoes in South America. Circles show the location of earthquakes and triangles show the location of volcanoes with distance from the trench.

In domains such as geosciences, students usually visualize the 3D structure of objects presented on 2D interfaces (e.g., desktop computers) and then extract 2D profiles from the 2D representation of the data. For instance, phenomena and observations related to plate tectonics are inherently three-dimensional, yet are often plotted on 2D maps. In introductory geoscience courses, students are often trained to visualize 3D data by learning how to read 2D maps and block diagrams. For instance, this method of representation makes it difficult for some students to visualize the depth, extent, and geometry of earthquakes as they have different levels of penetrative thinking abilities. A 3D representation of the data can aid in better understanding the extent, shape, and cross-sections of the data. As an example, **Figure 2** shows the cross-section of earthquakes and volcanoes across South America. Drawing a cross-section based on a 3D visualization of data can be much easier than seeing the 2D representation of data, imagining the 3D visualization, and then extracting the 2D profile.

2.3 Sense of Embodiment in The Context of Penetrative Thinking

This research examines whether penetrative thinking as a topic in spatial learning can benefit from embodied learning. We incorporate embodied interactions with the 3D visualization of the data (earthquakes, volcanoes, and plate boundaries) to enhance students' ability in visualizing penetrative views and better understand the cross-section or profile of the data in different regions around the world. To evaluate the role of bodily engagement through different interaction techniques introduced in **Section 2.1**, object manipulation and moving the user in space, in a penetrative thinking exercise, we compared the two design choices by providing two VLEs in the form of web-based desktop VR applications. These VLEs are designed to create an interactive environment to support penetrative thinking in an introductory physical geology course to facilitate visualization of the distribution and depth of earthquakes around the world. Full bodily engagement and a higher level of embodiment can be achieved in an immersive VR using Head-Mounted Displays (HMDs). In a web-based desktop VR application, a lower level of bodily engagement can be created through hand movements and the use of a device like a mouse or a keyboard.

In the first condition, where bodily engagement is induced through object manipulation, students do not actively move in the environment. They move and manipulate all the 3D objects together by dragging, rotating, or zooming in/out. This manipulation technique helps the students to get closer to a specific location along x,y,z direction, where they can observe a specific subduction zone. In this condition, students have complete control over manipulating all 3D objects at the same time. They can switch between different datasets but they cannot manipulate each object individually (i.e, individual earthquake locations or volcanoes). We refer to this visualization as *the drag and scroll* condition (**Supplementary Video S1**). This interaction technique is similar to what is experienced in conventional 3D editors or geoscience software programs such as ArcScene[4].

In the second condition, where the bodily engagement is induced through moving the user in space and creating a sense of locomotion, students rotate the viewpoint to the desired direction (along x,y,z axes) and move farther and closer to the 3D objects to inspect their spatial arrangement and their associated information. In this condition, the user can move in space and change the direction of the viewpoint in the virtual environment in a natural way (similar to what is experienced in conventional first-person camera views in games). In this condition, we manipulate the position and rotation of the first-person camera in VR to create a sense of egocentric movement in space. The first-person camera manipulation is designed based on the rotation of the camera using the mouse for determining the direction of the viewpoint and the arrow keys on the keyboard to translate in that direction. we refer to this condition as the *first-person* condition (see the **Supplementary Video S1**). This type of interaction technique in a web-based desktop VR is the closest type of simulation that we could create to induce the sense of locomotion compared to physical walking in an immersive VR using HMDs. Based on these definitions, the main difference between these two interaction techniques is the design choice of moving the 3D objects or moving the user.

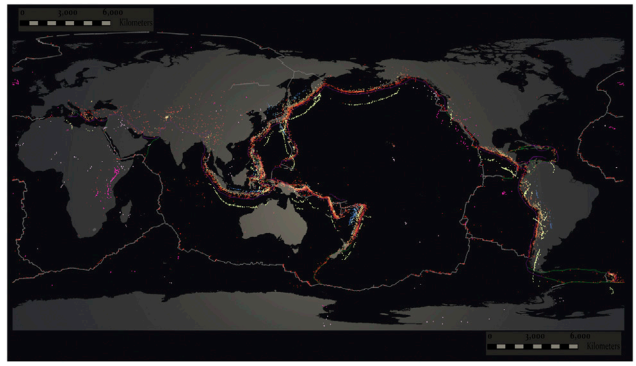

FIGURE 3 | Top-down view of the web-based desktop VR application showing the world map, plate boundaries, earthquakes and volcanoes. **Figure 5** shows the legend.

3 THE EXPERIMENT

This research examines the role of bodily engagement as an embodied affordance on users' learning experience and performance. To conduct this research, two types of interaction techniques have been defined that can affect bodily engagement and the associated sense of agency. At the time of epidemic of COVID-19, when the use of HMDs became limited for safety reasons, designing web-based desktop VR applications that are accessible via web browsers gave students the flexibility of going through the exercise at home on their personal computers. We designed two web-based VLEs to explore how the design choices of interaction techniques can affect bodily engagement, agency, learning experience, and performance. As a case study, we visualized 3D earthquake locations around the world representing the USGS Centennial Earthquake Catalog (Ritzwoller et al., 2002) and Holocene volcanoes (Venzke, 2013) in the context of plate boundaries (Coffin et al., 1997).

Figure 3 shows the top-down view of the web-based desktop VR applications and **Figure 4** shows an egocentric view. The design of each VLEs is the same in terms of data visualization. What makes the two different is how interaction with the datasets is realized, which can be shown in a recorded video but not in a figure.[5] The first VLE uses a mouse to drag, rotate and zoom in/out of the 3D visualization of the earthquakes and volcanoes. We refer to this visualization as the drag and scroll condition. The second VLE uses a mouse to define the direction of the viewpoint and the keyboard's arrow keys to translate in the environment. We name this 3D visualization as the first-person condition.

Considering these two experimental conditions, this study investigates the following hypotheses in two area of interests: learning experience and learning performance.

[4]https://desktop.arcgis.com/en/arcmap/latest/extensions/3d-analyst/choosing-the-3d-display-environment.htm

[5]Please refer to the video of the interaction techniques provided as the supplemental material

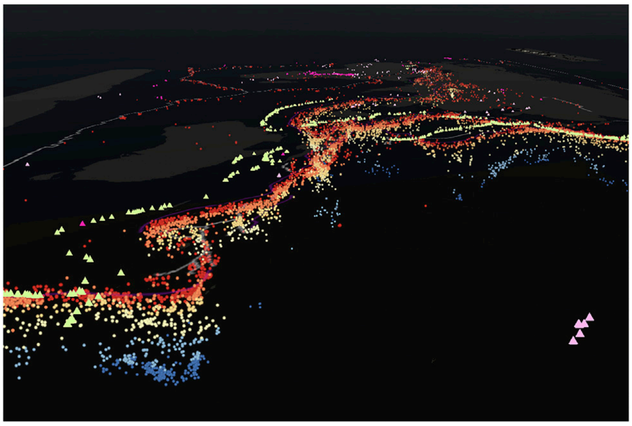

FIGURE 4 | Egocentric view of the USGS Centennial Earthquake Catalog and Holocene Volcanoes. **Figure 5** shows the legend.

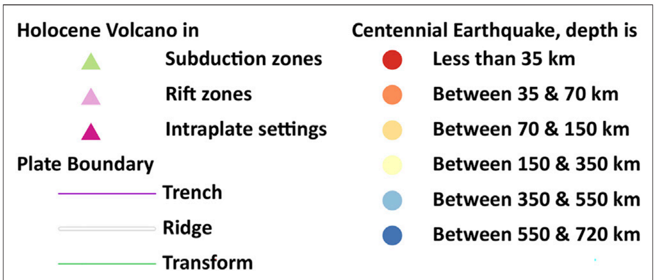

FIGURE 5 | Legend of the data visualized in the application, including plate boundaries, earthquakes and volcanoes.

Learning Experience:

H1. Students in the first-person condition experience a higher sense of presence.
H2. Students in the drag and scroll condition have higher control over the learning materials and as a result experience more agency.
H3. Students report a higher level of perceived learning in the drag and scroll condition.
H4. Students with a higher level of Visual Spatial imagery ability experience a higher sense of presence regardless of the condition.

Learning Performance:

H5. Students' learning performance with low knowledge of the field improves after going through the experiences regardless of the conditions.
H6. Students' level of control positively affects their learning performance regardless of the condition.
H7. Students with higher penetrative thinking ability show higher learning performance regardless of the condition.
H8. Students with lower penetrative thinking ability perform better in the first-person condition.

3.1 System Design

The data used to realize the visualizations in both conditions is the USGS Centennial Earthquake Catalog, which is a global catalog of well-located earthquakes from 1900 to 2008 that allows for the investigation of the depth and lateral extent of seismicity at plate boundaries (Coffin et al., 1997). To complement the earthquake locations and further connect the exercise to plate tectonics and plate boundary zones, maps of the current plate boundaries and the location of Holocene (i.e., < 10,000 years) volcanoes are also provided. **Figure 5** shows the information provided in both conditions: 1) the three main plate boundary types; 2) horizontal scale in km; 3) the depth of the earthquakes: depth is less than 35 km; depth is between 35 and 70 km; depth is between 70 and 150 km; depth is between 150 and 350 km; depth is between 350 and 550 km; depth is between 550 and 720 km; 4) volcanoes: in subduction zones, in rift zones, and intraplate settings. The original format of the USGS Centennial Earthquake Catalog was a text file and for the Holocene Volcano, the original format was an Excel XML, both containing several values including X, Y, Z. The coordinates stored in the tables were imported into ArcGIS Pro[6] as XY point data using the XY Table to Point tool.

The shapefiles were imported into Blender (Community, 2018) using a Blender importer called BlenderGIS[7]. Then they were imported to Unity3D®[8] as FBX files. The earthquakes and volcanoes were visualized in the form of point clouds and were properly georeferenced. To overcome the performance limitation of rendering a large dataset (a total of 13,077 points for earthquakes) in VR, we used the particle system of Unity3D to generate points to have a more efficient and performant experience. Plate boundaries were visualized in the form of lines overlaid on the world map. Using these datasets, students can examine different subduction zone plate tectonics in terms of the locations and depths of the earthquakes.

The two different interaction techniques (one per condition) with the datasets were implemented in Unity3D. In both conditions, the users can switch between the earthquake and volcanoes datasets or enable both at the same time. Furthermore, they can access the label and other information of the data by opening showing/hiding a legend of the dataset. There is a scale bar next to the map to help users with the perception of distances. In the drag and scroll condition, the view of the users (i.e., the camera) orbits around a pivot point (starting at the center of the scene) using a common drag and movement functionality with the right mouse button, allowing the user to rotate the viewpoint. In addition, the pivot point can be moved within the 3D space of the scene along the X, Y, and Z axes using the drag and movement functionality with the left mouse button. Doing so would enable the users to move along these axes, and consequently orbit around the new pivot position. In the first-person condition, the users will use a combination of mouse and keyboard to perform a smooth translation along the X, Y, and Z axes using the WASD (or arrow) keys on the keyboard, while changing the

[6]https://www.esri.com/en-us/arcgis/products/arcgis-pro/resources
[7]https://github.com/domlysz/BlenderGIS
[8]https://unity.com

direction of the movement based on the rotation of the camera using the mouse (i.e., steering which direction to move to with the mouse while the force is applied to that direction via the keyboard keys). The locomotion techniques in the conditions are very similar in nature (virtual travel and view point manipulation), but the two conditions are different in the mechanics of interaction used for locomotion. The drag and scroll condition simulates the interaction mechanics in software like ArcGIS, and the first-person condition simulates the interaction mechanics found in typical first-person shooter game.

3.2 Participants

236 students from two separate sections of an introductory physical geology course were invited to participate in this study in the Fall of 2020. The experience was embedded into the course as a lab assignment. Using a web-page, students selected whether they would like to take part in the research or only do the exercise as a lab assignment. From the 177 students who agreed to participate in the study, 96 students were randomly assigned to the drag and scroll condition and 81 students to the first-person condition. The section enrollment of participants was anonymized during the condition assignment to control for the environmental factors. All participants were compensated with extra course credit for their participation. 29.94% of the participants were female, 67.79% male, and less than 3% declared were non-binary or gender-nonconforming. The average age of the students was 19.45, with a maximum age of 21 and a standard deviation of 0.83. Also, 73.44% of the participants were majoring in Engineering.

3.3 Measures and Tests

To measure learning experience and knowledge gain, two types of questions were used in this study: 1) standardized measures, and 2) knowledge tests. Several existing standardized measures were incorporated into the pre-, and post-study questionnaires. Except for the demographic and background questions, all measures were of the type Likert-scale (ranging from 1 to 5 with 5 being the most positive), open-ended or multiple choice.

The pre-study questionnaire was comprised of the following measures:

- Demographics and background-related questions about gender, age, major and minor fields of study, and the year of study.
- A self-report measure of individual differences in terms of visual imagery: using the Visual Spatial Imagery (VSI) from MEC Spatial Presence questionnaire (Vorderer et al., 2004), with each item measured on a 1 to 5 Likert-scale. VSI is one of the spatial abilities that measures the ability to create clear spatial images and later access them from memory. People with higher VSI ability find it easier to access those spatial images from their memory (Wirth et al., 2007).

The post-study questionnaire was used to assess the learning experience of participants in light of the sense of presence and the sense of agency. Furthermore, the perceived learning experience of participants was measured.

- For measuring the sense of presence, we used the 6-item metric of Spatial Situation Model (SSM) from the MEC Spatial Presence Questionnaire (Vorderer et al., 2004). According to Wirth et al. (2007), a sense of presence can be built based on the Spatial Situation Model (SSM).
- For measuring the sense of agency, we used a combination of measures including Possible Actions from the MEC Spatial Presence Questionnaire (Vorderer et al., 2004) and measures suggested by Lee et al. (2010) including immediacy of control, perceived ease of use, and control and active learning.
- To measure perceived learning experience, we used three measures by Lee et al. (2010): reflective thinking, perceived learning effectiveness, and satisfaction. Perceived learning gives us feedback on the learning experience of students.
- Two open-ended questions were used to capture the general impression of participants about what they would change in the experiment and the advantages and disadvantages of this method of learning compared to classical teaching methods in classrooms.

For the knowledge tests, a pre-study and a post-study test were designed. Besides, a test that measured the participants' mental slicing and penetrative thinking ability was used:

- The pre-study knowledge test contained six multiple-choice questions that tested students' pre-knowledge of subduction zones and plate boundaries before going through the main experience.
- In the post-study knowledge test, seven multiple-choice questions were asked from the students to test their knowledge of the subject based on their penetrative thinking ability. In the pilot study (Bagher et al., 2020), we asked the students to draw by hand cross-sections plotting the depth of the earthquakes with distance from a subduction zone trench for segments of South America and Japan. Drawing a cross-section is a straightforward technique to test the students' penetrative thinking ability in the field. In this research, due to remote participation, we could not include the same exercise. Therefore, we curated questions that not only test students' knowledge of the subduction zones but test their penetrative thinking ability in the context of earthquake depth and distribution. For instance, we asked the student "Below are cross-sections of seismicity versus depth for four different subduction zones. Which cross-section is most similar to the South America subduction zone?". The students had to use their VSI and penetrative thinking abilities to recall the cross-section of the South America subduction zone in their observation and choose one plot from multiple choices.
- The Geologic Block Cross-sectioning Test (GBCT) (Ormand et al., 2014) contains sixteen multiple-choice questions assessing the students' ability to understand three-dimensional relationships by determining the correct vertical cross-section from a geologic block diagram.

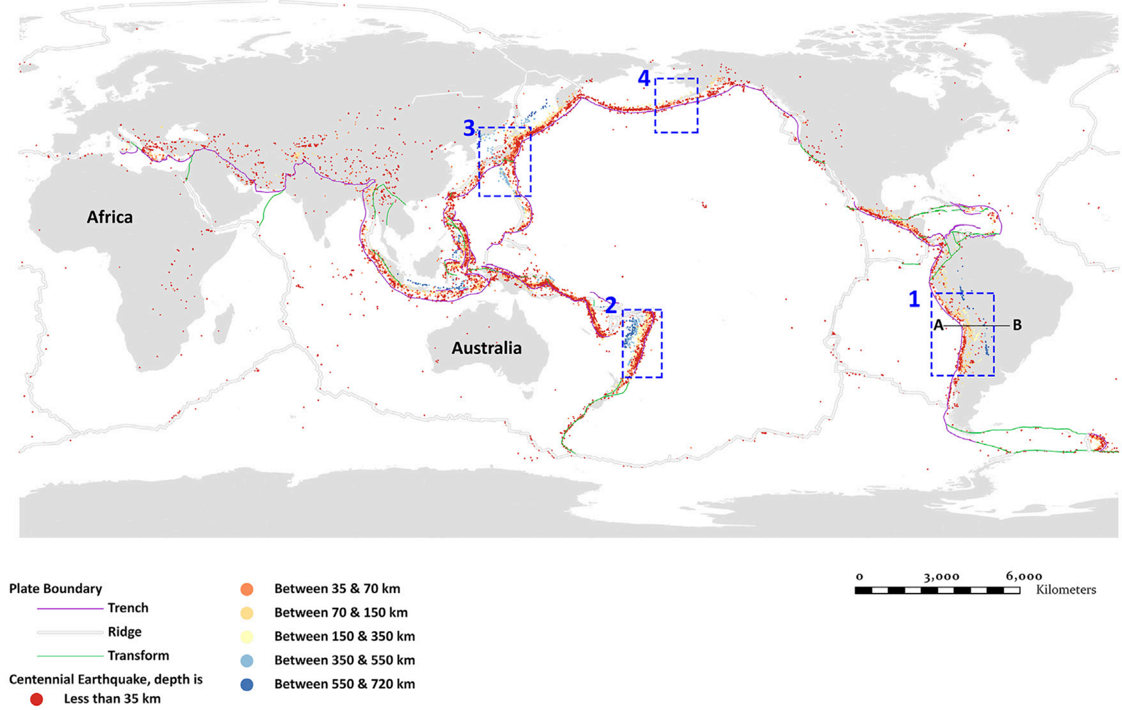

FIGURE 6 | Area of interests for the virtual experience. Students were asked to focuse on these eareas during the virtual experience.

3.4 Procedure

In both conditions, students filled out the pre-study questionnaire and then answered the pre-study knowledge test to establish their prior knowledge about the learning topic. Then, they were given information on the types of datasets they were going to explore in the VR experience and instructions on what areas to focus on. **Figure 6** shows the area of interests including boxes 1–4 and cross-section A-B.

Region 1: South America
Region 2: Tonga-Kermadec
Region 3: Japan
Region 4: Eastern Alaska
Cross-section A-B: A cross-section across South American convergent margin.

Students were asked to explore and pay attention to the distribution of the earthquakes and volcanoes, and the depth range of the earthquakes in these regions while reflecting on the following questions: What do you observe with respect to these different subduction zones? Are the geometries of the subducting oceanic lithosphere the same (i.e., the distribution and geometry of the earthquakes) or are they different? Now, look specifically at the western margin of the South American Plate (Region 1). Is the Wadati-Benioff zone (i.e., the zone of seismicity that defines the subducting plate) the same north to south along the margin? They were informed that after the experience, they will be asked to answer several questions about these regions and the cross-section. In both conditions, they were given 15 min to explore the datasets and memorize the distribution of earthquakes in the defined regions. A two-dimensional guide map on the lower right side of the screen showed the position and the direction of the user in the world map. A timer on the upper left side reminded them of the remaining time (**Figure 7**). In both conditions, students could hide/show legend and instructions.

After the experience, students first answered the post-study questionnaire, and then the penetrative thinking ability test. Finally, they answered the post-study knowledge test. Placing the post-study knowledge test at the end introduced a period between the experience and the post-study knowledge test. This way, we could test the effect of various embodied interactions on knowledge retention. The session, from start to end, took around 40 min.

3.5 Analysis

For the learning experience assessment, we first identified the outliers using the Interquartile Range (IQR) method and carefully checked the dataset for removing any outliers. Then, we used Welch's two-sample t-tests to compare the first-person condition with the drag and scroll condition based on the learning experience measures. For the learning performance measures, when Z-scores of the pre-, and post-study knowledge tests were compared regardless of the condition, Welch's two-sample t-test was calculated. When we compared the post-study grades among the conditions, since the grades were ranked data, the Wilcoxon signed-rank test was used. To predict students' sense of presence based on Visual Spatial Imagery and post-study grades based on their penetrative thinking ability, regression equations were

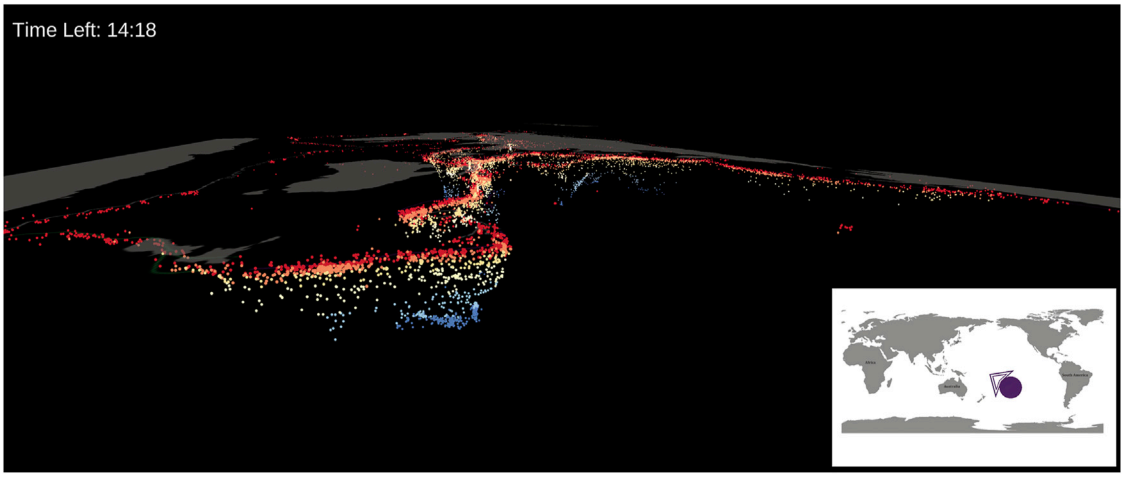

FIGURE 7 | Guide map and the time counter to help the students keep track of time and navigate in the learning environment.

calculated. As the number of participants in the two groups was different, Hedges' g (Hedges and Olkin, 2014) was calculated instead of Cohen's d for the calculation of effect size. A qualitative analysis of the two open-ended questions was performed to gain a better understanding of the participants' opinions and experiences. Based on the approach proposed by Schreier (2012), two independent coders went over the responses of participants and inductively generated codes that would capture their content. Followed by consensus meetings, the codes were then grouped or rearranged into the final schema. Inter-rater reliability tests based on Cohen's Kappa were then calculated for the finalized results.

4 RESULTS

4.1 Learning Experience Assessment

Table 1 presents an overview of the mean, standard deviation, *p*-value, and effect size of the experience measures in the drag and scroll and the first-person conditions. As mentioned in the measures section, we measured the sense of presence, sense of agency, and perceived learning experience. There was no significant difference between the two conditions in terms of the sense of presence. Therefore, the first hypothesis (students in the first-person condition experience a higher sense of presence) is rejected. In terms of sense of agency, we measured possible actions, immediacy of control, perceived ease of use, and control and active learning, introduced in Section C. There is a significant group difference in the ease of use scores between the first-person (M = 3.14, SD = 0.63) and the drag and scroll (M = 3.32, SD = 0.50) conditions in favor of the drag and scroll condition, [t (153.12) = −1.98, p = 0.04]. The immediacy of control measures the students' agency to change the view position and manipulate spatial objects. The difference for immediacy of control is very close to significant [t (174) = −1.77, p = 0.07] in favor of the drag and scroll condition (M = 4.07, SD = 0.88). We could not find any significant difference between the two conditions in terms of possible actions, and control and active learning. Based on these results, we have found some evidence in favor of the second hypothesis: students in the drag and scroll condition have higher control over the learning materials and as a result experience more agency. However, we could not find significant differences in all measures related to this affordance and as a result, we cannot conclude that the second hypothesis can entirely be accepted. In terms of perceived learning, students in the drag and scroll condition (M = 3.41, SD = 0.56) were significantly more satisfied [t (146) = 1.76, p = 0.04] than in the first-person condition (M = 3.20, SD = 0.75). We could not find any significant difference between the conditions in terms of reflective thinking and perceived learning effectiveness. Therefore, the only evidence that we could find in favor of the third hypothesis (students report a higher level of perceived learning in the drag and scroll condition) was satisfaction. Subsequently, we cannot conclude that the third hypothesis can be entirely accepted.

To conclude briefly, based on the discussed results, students in the drag and scroll condition had a better learning experience in terms of ease of use, immediacy of control, and satisfaction.

A simple linear regression was calculated to predict the effect of Visual Spatial Imagery (VSI) as a spatial ability on the sense of presence (SSM). Independent of the condition, a significant regression equation was found [F (1,175) = 53.04, p < 0.001] with an adjusted R^2 of 0.228. Students' sense of presence has increased by 0.64 for each unit of VSI. Therefore, hypothesis 4 can be accepted: students with a higher level of VSI experience a higher sense of presence. **Figure 8** shows that in both the drag and scroll and the first-person conditions, the level of presence is dependent on the VSI spatial ability. A significant regression equation was found for the first-person condition [F (1,79) = 28.64, p < 0.001] with an adjusted R^2 of 0.256. Students' sense of presence has increased by 0.66 for each unit of VSI. For the drag and scroll condition, the significant regression equation is [F (1,94) = 22.83, p < 0.001] with an adjusted R^2 of 0.186. Students' sense of presence has increased by 0.61 for each unit of VSI.

TABLE 1 | Overview of the learning experience measures.

	Measures	Conditions	M	SD	p	Effect size
Sense of presence	SSM	Drag and scroll	3.669	0.73	0.506	0.002
		First-person	3.667	0.85		
Sense of agency	Possible actions	Drag and scroll	3.50	0.70	0.2	0.12
		First-person	3.40	0.88		
	Ease of use	Drag and scroll	3.32	0.50	0.048*	0.31
		First-person	3.14	0.63		
	Immediacy of control	Drag and scroll	4.07	0.88	0.07	0.21
		First-person	3.84	0.87		
	Control and active learning	Drag and scroll	3.83	1.01	0.27	0.133
		First-person	3.96	0.93		
Perceived learning	Reflective thinking	Drag and scroll	3.64	0.73	0.09	0.192
		First-person	3.49	0.83		
	Perceived learning effectiveness	Drag and scroll	3.61	0.63	0.13	0.18
		First-person	3.48	0.79		
	Satisfaction	Drag and scroll	3.41	0.56	0.04*	0.32
		First-person	3.20	0.75		

4.2 Learning Performance Assessment

Before going through the experience, students answered six questions about subduction zones to test their knowledge of the subject in terms of the ability to understand the extent and geometry of the subduction zones based on their interpretation of the earthquakes, volcanoes, and plate boundaries. The total possible score was 8; The result of the test indicated an average score of 3.21 (SD = 1.24) with a minimum score of 1 and a maximum score of 7. 56.49% of the students who attended the study obtained a score that is less than the average score. This indicates that 56.49% of the students who attended the study had lower knowledge of the field compared to average performance. The post-study knowledge test contained seven questions with a total possible score of 14. The post-study knowledge test examined the same knowledge concepts with different types of questions to evaluate whether students' understanding of the subject has improved after going through the experience. The result of the test indicated an average score of 7.9 (SD = 2.32) with a minimum score of 3 and a maximum score of 13.

Comparing the Z-scores of the pre-, and post-study knowledge tests, regardless of the condition, shows that students' performance has improved by 0.05. However, the difference is not statistically significant: [t (176) = 0.55, p = 0.58]. We were under the impression that we can detect the presence or absence of students' knowledge gain by studying the whole sample size. However, students with higher prior knowledge have a different level of improvement than students with lower knowledge of the field. Subsequently, we decided to analyze the learning performance of students with low prior knowledge of the subject compared to the average performance (pre-test Z-score ≤ 0). Based on our analysis, the performance of students with low prior knowledge of the field improved significantly regardless of the conditions: [t (167) = −5.86, p < 0.001]. For the drag and scroll condition [t (52) = −3.34, p < 0.001] and for the first-person condition, [t (46) = −5.41, p < 0.001]. Therefore, hypothesis 5 is accepted: both conditions have a significantly positive effect on students with low prior knowledge of the subject and the exposure to the VLEs improved their learning performance in terms of understanding earthquakes' distribution and depth. In other words, when students with low prior knowledge of the field were exposed to the 3D representation of the epicenters of earthquakes from the USGS Centennial Earthquake catalog and locations of Holocene volcanoes, they understood the locations, depth, and geometry of the earthquakes in subduction zones better in different regions through 3D visualization. Yet, we could not find a significant difference between the conditions in terms of knowledge gain in students with low prior knowledge of the field: [t (97.2) = 0.94, p = 0.34].

We also analyzed the impact of the immediacy of control as one of the important measures of the sense of agency on students' learning performance (post-study grades). The students' post-study grades were dependent on their evaluation of the immediacy of control in both conditions. In both conditions, the higher a student felt to be in control, the higher their post-study grades were (**Figure 9**). In the drag and scroll condition, a significant non-linear regression equation was found [F (1,92) = 3.406, p = 0.02] with an adjusted R^2 of 0.07. In the first-person condition, a significant regression equation was found [F (1,77) = 3.007, p = 0.03] with an adjusted R^2 of 0.069. Although the adjusted R^2 for both equations are incredibly low and show that the immediacy of control is not a strong contributing factor, it is worth mentioning that there is a significant correlation. Respectively, hypothesis 6 is accepted: a higher level of control positively affects students' learning performance.

To measure students' penetrative thinking ability, students were asked to take the Geologic Block Cross-sectioning Test (GBCT) (Ormand et al., 2014). A simple linear regression was calculated to predict the result of the post-study grades based on the GBCT score (penetrative thinking ability). Independent of the condition, a significant regression equation was found [F (1,175) = 21.87, p < 0.001] with an adjusted R^2 of 0.106. Therefore,

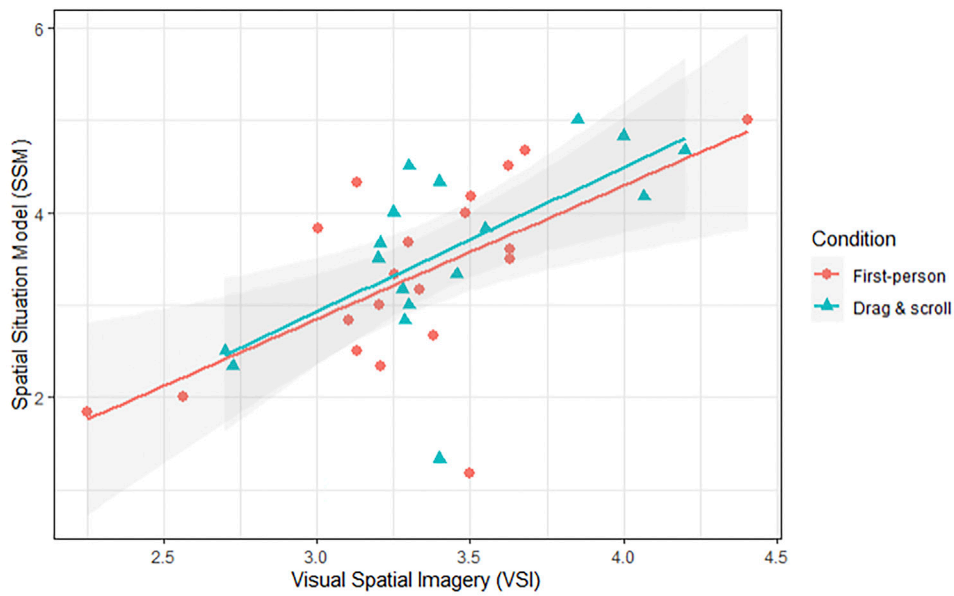

FIGURE 8 | The plot of VSI and SSM for each condition.

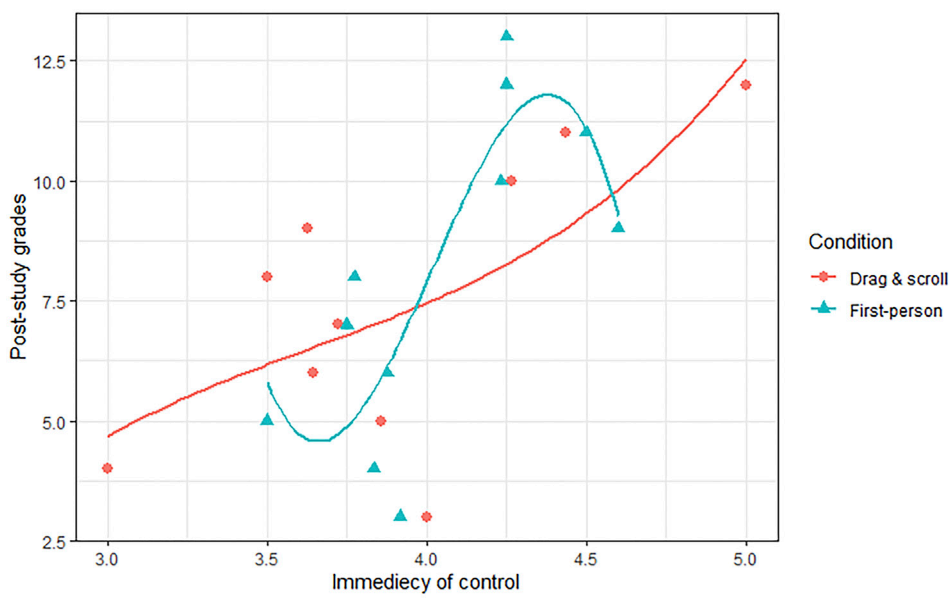

FIGURE 9 | The plot of post-study grades and the immediacy of control for each condition.

hypothesis 7 about learning performance is accepted: students with higher penetrative thinking ability show higher learning performance. This shows that for students who understand the spatial relations between the objects, this penetrative thinking ability enables them to understand the location, direction, and shape of earthquake events around the world better. In terms of students with lower penetrative thinking ability (hypothesis 8), there is a significant difference between the post-study knowledge grades of the first-person condition ($M = 7.84$, $SD = 2.07$) and the drag and scroll condition ($M = 6.63$, $SD = 2.5$) in favor of the first-person condition, [$t\,(67) = 2.36$, $p = 0.02$]. We can conclude that the first-person condition with the freedom of moving in space and inspecting earthquake locations by moving closer to the objects in a first-person view has a positive effect on students with a low penetrative thinking ability. Therefore, hypothesis 6 is accepted: students with lower penetrative thinking ability perform better in the first-person condition. Interestingly, in the drag and scroll condition, there is a significant difference

TABLE 2 | Summary of the structured content analysis.

Question	Code	% Participants in the drag and scroll condition	% Participants in the first-person condition	Cohen's kappa
Q1	Add more comprehensive information	18.7	22.2	0.859
	Accessibility to learning objectives in VR	12.5	16	0.883
	Difficulty navigating in the space	0.09	0.09	0.965
	Requesting more interaction	0.07	0.06	0.9
	Make the experience more visually appealing	0.10	0.02	0.9
	Negative learning experience	0.06	0.03	0.943
	Different movement mechanics using the keyboard	0	0.08	0.919
	Difficulty running the app	0.04	0.03	1
	The experience was too long	0.03	0	1
	Improve navigation with the mousepad	0	0.01	1
	Preferring drag and scroll over moving in space using the keyboard	0	0.01	1
	Clear representation of the distance between objects	0	0.01	1
	Have a zoom function with the real images of locations	0.01	0.01	1
	Using a different method for switching between datasets	0.0	0.01	1
	Prefer HMD over the web application	0.01	0	1
	Show legend at all times	0.01	0	1
	Suggest quick jump navigation technique	0.01	0	1
Q2	The experience has advantages over classical teaching methods	54.16	40.7	0.977
	The 3D representation and interactive features improved understanding of the concepts	37.5	38.2	0.911
	The experience has no advantage over classical teaching methods	0.09	19.7	1
	Learning at your own pace	11.04	0.02	0.907
	Prefer the classical teaching method	0.04	0.07	0.943
	The experience lacked active QA with the instructor	0.06	0.01	1
	The experience was easier	0	0.02	1
	Superior to other remote learning approaches	0	0.01	1

between the pre-, and post-study grades (Z-scores) of students with a low penetrative thinking ability (mean of the differences = 0.45) [t (37) = 2.11, p = 0.04]. Although students with a low penetrative thinking ability in the drag and scroll condition had lower post-study grades compared to the first-person condition they had a significant improvement from their pre-study grades. This result indicates that even though the drag and scroll condition is not as effective as the first-person condition in terms of the knowledge gained in students with a low penetrative ability, it still is an effective medium and has improved students' knowledge gain after being exposed to the VR experience.

4.3 Qualitative Analysis of the Open-Ended Feedback of The Experience

Two open-ended questions were asked from the participants about their experiences as part of the post-study questionnaire:

Q1: If you could have changed something in the experience what would it have been and why?
Q2: If any, did this current method of instruction have advantages over classical methods of teachings used in classrooms?

Along with the quantitative analysis, the conducted qualitative analysis provides insights into the experiences of users after going through each condition. The extracted codes, capturing the content of the comments by participants, the percentage of participants talking about a code, and Cohen's Kappa inter-rater reliability coefficient are reported in **Table 2**. Some of the codes are generally applicable to the experience regardless of the conditions and some are specific to the design choices based on the condition.

For the first question, the most frequent code was requiring more comprehensive information. Examples of this code include requesting an interactive legend (i.e., audio feedback), detailed description of features, and adding more features (i.e., mountains or continent names). In terms of accessibility to learning objectives in VR, before the experience, instructions and learning objectives were given to the students, including the highlighted areas to focus on and questions to have in mind while exploring the datasets. However, many students felt the need to see these learning objectives in the VR experience, being able to turn on and off the highlighted areas, and receiving more educational explanations of various subduction zones in the form of audio or text instead of self-exploration. Some students felt that there is no need for a change in the application whereas others mentioned difficulty in navigating in the space, negative learning experience, and difficulties running the app. In the first-person condition, some suggested different movement mechanics be designed to improve the experience. They suggested that instead of using the mouse as defining the direction of movement, two keys on the keyboard should allow for up and down movement. No one in the first-person condition complained that the experience was too long while three people in the drag and scroll condition complained about the length of the experience. Since the method of interaction in the first-person condition was new and students were not familiar with this method of movement in space, they might have used a considerable amount of time learning how to navigate in space and did not feel the time passing. Whereas the method of interaction in the drag and scroll condition is similar to geoscience software programs that many are familiar with.

In response to the second question, almost half of the students found this method of teaching superior to the classical methods of teaching. The advantages counted for this experience included learning at your own pace, being easier, and indicating that the 3D representation and interactive features improved their understanding of the concepts. 11.04% of the students in the drag and scroll condition declared that this method helped them to learn at their own pace, while only 0.02% in the first-person condition felt that way. As mentioned in the analysis of the first question, students in the first-person condition might have used a considerable amount of time learning how to navigate in space and that might have affected their learning pace. On the other hand, 19.7% of the students in the first-person condition found this type of experience to have no advantages over classical methods of teaching while only 0.09% in the drag and scroll condition felt that way. This indicates that although 40.7% of the students in the first-person condition found this method advantageous, 19.7% disagreed. One of the negative feedback about this method of teaching was the absence of active Q&A with the instructor while learning.

Overall, the insights from the first question show that students enrolled in the physical geology course are not used to memorization tasks, they typically would plot the locations and depths of the earthquakes by directly observing the data. In the exercise we designed, they first observed the data and then recalled the cross-sections based on memorization and memory trace. The second question gives insight that half of the students are open to technology-integrated teaching methods. Perhaps by improving their experience regarding the issues mentioned in the coding of the first question, more students might be open to this method of teaching.

5 DISCUSSION

This study investigated the impact of bodily engagement on the learning experience and performance in the context of penetrative thinking in a critical 3D task in geosciences education: understanding the cross-section of the depth and geometry of earthquakes with distance from the trench. Since we have used the same platform (web-based desktop VR) for the design of VLEs, this study is not focusing on the effect of different mediums or degrees of embodiment on learning but the impact of interaction techniques on learning experience and performance.

5.1 The Effect of Bodily Engagement on Learning Experience

Our quantitative evaluation of the learning experience utilized established self-reported measures. We were anticipating a significant difference in the sense of presence between the two conditions. Although the sense of presence is not significantly different among the two conditions, we found that students with higher Visual Spatial Imagery (VSI) ability experience a higher sense of presence in both conditions. In terms of perceived learning, we found that students are significantly more satisfied with the drag and scroll condition but we could not find any difference in other measures related to perceived learning. Concerning the sense of agency, students reported that the drag and scroll condition is significantly easier to use than the first-person condition. They also found the drag and scroll condition to have a higher level of immediacy of control compared to the first-person condition. This evaluation indicates that students are more comfortable and familiar with the interaction method data manipulation which is dragging, rotating, and zooming in/out of the 3D data. This made us curious to see if declaring the drag and scroll condition as an easier interaction technique would translate into superior knowledge gain as well. The results of the structured content analysis show that almost the same percentage of students in both conditions felt that the 3D representation and the method of interaction have improved their understanding of the subject.

5.2 The Effect of Bodily Engagement on Learning Performance

Overall, all students gained some knowledge by going through the experience but we aimed to investigate the impact of interaction techniques on knowledge gain for students with low prior knowledge of the field. Our analysis showed that knowledge gain in students with low prior knowledge of the field improved significantly after going through the virtual experience in both conditions. We also found that when students felt more in control, in both conditions, they significantly performed better in terms of knowledge gain. This demonstrates that having control can be a contributing factor in knowledge gain. This shows that some students are more comfortable with moving in the three-dimensional environment and inspect objects based on changing their viewpoint whereas some students are more comfortable with data manipulation. With this result in mind, we looked into the penetrative thinking ability of the students to find out whether it would play a role in knowledge gain in different conditions. In the next section, we discuss our findings regarding students with lower penetrative thinking ability.

5.3 The Overall Effect of Bodily Engagement on Students With Lower Penetrative Thinking Ability

Weise et al. (2019) advise that the characteristics of the users should be considered in choosing an interaction technique. They suggest that users' abilities can affect the performance and usability of the interaction technique. In this study, We used the Geologic Block Cross-sectioning Test (GBCT) to evaluate students' penetrative thinking ability. We assessed whether this ability might affect their performance using either of the interaction techniques. Regardless of the conditions, we observed that the higher the penetrative thinking ability of the students, the higher the knowledge gain was. We hypothesized that students with higher spatial ability would better understand spatial relations of 3D objects and would perform better in either condition. One goal of designing interactive and embodied VLEs in 3D is to help students with lower spatial abilities, to help them visualize data in 3D, and better understand spatial relations between 3D objects. We found that students with a lower penetrative thinking ability benefited more from the interaction of the first-person condition. They had a significantly higher knowledge gain than students with a lower penetrative thinking ability in the drag and scroll condition. This result indicates that students with lower penetrative thinking ability benefit from active movement in space that facilitates adjusting their viewpoints. In other words, manipulating objects and trying to rotate them to get the desired viewpoint might be complex for students with lower penetrative thinking ability than naturally moving in space. Even though students with a lower penetrative thinking ability performed significantly higher in the first-person condition in terms of knowledge gain, students with a low penetrative thinking ability in the drag and scroll condition improved significantly compared to their pre-test Z-Score. This result suggests that even though the drag and scroll condition is not as ideal as the first-person condition for these students in terms of post-study knowledge gain, being exposed to a 3D representation of the data and interacting with the data would improve students' penetrative view and result in a higher understanding the locations and depths of earthquakes when they have low penetrative thinking ability.

6 CONCLUSION, LIMITATIONS, AND FUTURE WORK

In this article, we explored students' penetrative thinking ability to interpret subduction zone plate tectonics from observations of the locations and depths of earthquakes. We argued that embodied learning could promote students' learning experience and performance in visual-spatial thinking tasks such as penetrative thinking. To examine the role of bodily engagement as an embodied affordance on students' learning experience and performance in an introductory physical geology course, we designed two VLEs based on two different interaction techniques: 1) object manipulation (drag and scroll) and 2) moving the user in space (first-person). Analyses of the data concerning learning experience and performance provided us with insights into students' perception of learning and the actual performance. Overall, we argue that both interaction techniques have pros and cons regarding learning experience and performance. The goal of the VLE and the students' spatial ability can further define which condition is a more suitable choice for teaching earthquake locations and depths.

One of the limitations of this study is the gender composition consisting of primarily male participants. Although our focus has not been the gender differences in spatial abilities, we are aware that there are conflicting studies regarding the differences in spatial abilities among male and female participants Yuan et al. (2019). Unfortunately, most studies focusing on spatial abilities compare the performance of male and female participants and would not include non-binary participants. Another limitation is that although our population is from different fields and backgrounds, they have been examined in the context of geosciences. In future studies, we plan to investigate the role of bodily engagement in other courses concerning visual-spatial learning. Furthermore, to measure the effect of bodily engagement on knowledge retention, we had to ask the students to answer the post-study knowledge test in a couple of hours to a day. However, due to time constraints during data collection, we could only delay answering the post-study knowledge test by approximately 15 min. We introduced this period between the experience and the post-study knowledge test by placing the post-study knowledge test at the end of the post-study survey. Another limitation of this research pertains to the setup of the experiment. Like most research in this domain, our conclusions are based on a single exposure to the VLE. Using a longitudinal study with multiple exposures, the observed effects of bodily engagement between the used conditions could be either amplified or diminished. Therefore, we will perform a longitudinal study over several weeks to further explore the lasting effect of different interaction techniques in future research.

Due to the COVID-19 pandemic, we could not compare the effects of different mediums (IVR vs web-based desktop VR) on bodily engagement and embodied learning. Therefore, as part of the future work, we are devising methods for sending Oculus Quest headsets to the students for remote VR data collection. We opt to investigate the effect of a higher level of bodily engagement in IVR on learning. Furthermore, although we designed this experiment with the utmost care, we plan to implement improvements for future studies. For instance, for the design of the VLEs, we did not include audio feedback for gaining information on earthquake depth or types of volcanoes. This proved to be a sought-after feature by the students, and as such, will be included in future versions of the tool. Future studies will also aim to understand why students reported the drag and scroll condition to be easier to use. We hypothesize that familiarity with this method of interaction due to prior experiences with geological software might be a key predictor. However, it is also pertinent to investigate whether the use of Quest controllers for object manipulation in an immersive VR while physically walking in the environment is considered easier than object manipulation using a drag and scroll technique (web-based desktop VR). Furthermore, comparing an IVR with web-based desktop VR, we plan to investigate the level of control experienced by the students in each condition to explore how much sense of agency they would experience.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the IRB Program (Office of Research Protection) at Penn State University Study ID: STUDY00008293. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MB: First Author, conducted research, designed the VR experiment and the survey, carried out the analysis, and writing PS: designed the VR experiment, carried out the qualitative analysis, and writing JW: research design, quantitative analysis, and writing PLF: advisor on research design and implementation, writing AK: advisor on research design and implementation, writing.

FUNDING

This work was supported through a Penn State Strategic Planning award (proposal #1685_TE_Cycle2).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frvir.2021.695312/full#supplementary-material

REFERENCES

Ai-Lim Lee, E., Wong, K. W., and Fung, C. C. (2010). How Does Desktop Virtual Reality Enhance Learning Outcomes? a Structural Equation Modeling Approach. *Comput. Edu.* 55, 1424–1442. doi:10.1016/j.compedu.2010.06.006

Bagher, M. M. (2020). “Immersive Vr and Embodied Learning: The Role of Embodied Affordances in the Long-Term Retention of Semantic Knowledge,” in 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW) (IEEE), 537–538. doi:10.1109/vrw50115.2020.00120

Bagher, M. M., Sajjadi, P., Carr, J., La Femina, P., and Klippel, A. (2020). “Fostering Penetrative Thinking in Geosciences through Immersive Experiences: A Case Study in Visualizing Earthquake Locations in 3d,” in 2020 6th International Conference of the Immersive Learning Research Network (iLRN) (IEEE), 132–139. doi:10.23919/ilrn47897.2020.9155123

Bailey, J., Bailenson, J. N., Won, A. S., Flora, J., and Armel, K. C. (2012). “Presence and Memory: Immersive Virtual Reality Effects on Cued Recall,” in Proceedings of the International Society for Presence Research Annual Conference (Citeseer) (IEEE), 24–26.

Barsalou, L. W. (2008). Grounded Cognition. *Annu. Rev. Psychol.* 59, 617–645. doi:10.1146/annurev.psych.59.103006.093639

Barsalou, L. W. (1999). Perceptual Symbol Systems. *Behav. Brain Sci.* 22, 577–660. doi:10.1017/s0140525x99002149

Biocca, F. (1999). The Cyborg’s Dilemma. *Hum. Factors Inf. Tech.* 13, 113–144. doi:10.1016/s0923-8433(99)80011-2

Bulu, S. T. (2012). Place Presence, Social Presence, Co-presence, and Satisfaction in Virtual Worlds. *Comput. Edu.* 58, 154–161. doi:10.1016/j.compedu.2011.08.024

Clifton, P. G., Chang, J. S., Yeboah, G., Doucette, A., Chandrasekharan, S., Nitsche, M., et al. (2016). Design of Embodied Interfaces for Engaging Spatial Cognition. *Cogn. Res. Princ Implic.* 1, 24–15. doi:10.1186/s41235-016-0032-5

Coffin, M. F., Gahagan, L. M., and Lawver, L. A. (1997). *Present-day Plate Boundary Digital Data Compilation*. Austin, TX: Tech. rep., Institute for Geophysics.

Community, B. O. (2018). *Blender - a 3D Modelling and Rendering Package*. Amsterdam: Blender Foundation, Stichting Blender Foundation.

Czerwinski, E., Goodell, J., Sottilare, R., and Wagner, E. (2020). “Learning Engineering@ Scale,” in Proceedings of the Seventh ACM Conference on Learning@ Scale (ACM), 221–223. doi:10.1145/3386527.3405934

Dalgarno, B., Lee, M. J., Carlson, L., Gregory, S., and Tynan, B. (2011). An Australian and new zealand Scoping Study on the Use of 3d Immersive Virtual Worlds in Higher Education. *Australas. J. Educ. Tech.* 27, 1–15. doi:10.14742/ajet.978

Dalgarno, B., and Lee, M. J. W. (2010). What Are the Learning Affordances of 3-d Virtual Environments?. *Br. J. Educ. Tech.* 41, 10–32. doi:10.1111/j.1467-8535.2009.01038.x

Di Luca, M., Seifi, H., Egan, S., and Gonzalez-Franco, M. (2021). “Locomotion Vault: the Extra Mile in Analyzing Vr Locomotion Techniques,” in Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (IEEE), 1–10. doi:10.1145/3411764.3445319

Gonzalez-Franco, M., and Peck, T. C. (2018). Avatar Embodiment. Towards a Standardized Questionnaire. *Front. Robot. AI* 5, 74. doi:10.3389/frobt.2018.00074

Hannula, K. A. (2019). Do geology Field Courses Improve Penetrative Thinking? *J. Geosci. Edu.* 67, 143–160. doi:10.1080/10899995.2018.1548004

Hedges, L. V., and Olkin, I. (2014). *Statistical Methods for Meta-Analysis*. Academic Press.

Hegarty, M., Montello, D. R., Richardson, A. E., Ishikawa, T., and Lovelace, K. (2006). Spatial Abilities at Different Scales: Individual Differences in Aptitude-Test Performance and Spatial-Layout Learning. *Intelligence* 34, 151–176. doi:10.1016/j.intell.2005.09.005

Hostetter, A. B., and Alibali, M. W. (2008). Visible Embodiment: Gestures as Simulated Action. *Psychon. Bull. Rev.* 15, 495–514. doi:10.3758/pbr.15.3.495

Jerald, J. (2016). *The VR Book: Human-Centered Design for Virtual Reality*. Williston, VT: Morgan & Claypool.

Johnson-Glenberg, M. C., Bartolomea, H., and Kalina, E. (2021). Platform Is Not Destiny: Embodied Learning Effects Comparing 2d Desktop to 3d Virtual Reality Stem Experiences. *J. Comp. Assist. Learn.* 37, 1263–1284. doi:10.1111/jcal.12567

Johnson-Glenberg, M. C., Birchfield, D. A., Tolentino, L., and Koziupa, T. (2014). Collaborative Embodied Learning in Mixed Reality Motion-Capture Environments: Two Science Studies. *J. Educ. Psychol.* 106, 86–104. doi:10.1037/a0034008

Johnson-Glenberg, M. C. (2018). Immersive Vr and Education: Embodied Design Principles that Include Gesture and Hand Controls. *Front. Robot. AI* 5, 1–19. doi:10.3389/frobt.2018.00081

Johnson-Glenberg, M. C., Ly, V., Su, M., Zavala, R. N., Bartolomeo, H., and Kalina, E. (2020). “Embodied Agentic Stem Education: Effects of 3d Vr Compared to 2d Pc,” in 2020 6th International Conference of the Immersive Learning Research Network (iLRN) (IEEE), 24–30. doi:10.23919/ilrn47897.2020.9155155

Johnson-Glenberg, M. C., Megowan-Romanowicz, C., Birchfield, D. A., and Savio-Ramos, C. (2016). Effects of Embodied Learning and Digital Platform on the Retention of Physics Content: Centripetal Force. *Front. Psychol.* 7, 1819. doi:10.3389/fpsyg.2016.01819

Kelly, J. W., and McNamara, T. P. (2010). Reference Frames during the Acquisition and Development of Spatial Memories. *Cognition* 116, 409–420. doi:10.1016/j.cognition.2010.06.002

Kelly, J. W., and McNamara, T. P. (2008). Spatial Memories of Virtual Environments: How Egocentric Experience, Intrinsic Structure, and Extrinsic Structure Interact. *Psychon. Bull. Rev.* 15, 322–327. doi:10.3758/PBR.15.2.322

Kilteni, K., Groten, R., and Slater, M. (2012). The Sense of Embodiment in Virtual Reality. *Presence: Teleoperators and Virtual Environments* 21, 373–387. doi:10.1162/pres_a_00124

Klippel, A., Zhao, J., Jackson, K. L., La Femina, P., Stubbs, C., Wetzel, R., et al. (2019). Transforming Earth Science Education through Immersive

Experiences: Delivering on a Long Held Promise. *J. Educ. Comput. Res.* 57, 1745–1771. doi:10.1177/0735633119854025

Lages, W. S., and Bowman, D. A. (2018). Move the Object or Move Myself? Walking vs. Manipulation for the Examination of 3d Scientific Data. *Front. ICT* 5, 15. doi:10.3389/fict.2018.00015

Lee, K. M. (2004). Presence, Explicated. *Commun. Theor.* 14, 27–50. doi:10.1111/j.1468-2885.2004.tb00302.x

Legault, J., Zhao, J., Chi, Y.-A., Chen, W., Klippel, A., and Li, P. (2019). Immersive Virtual Reality as an Effective Tool for Second Language Vocabulary Learning. *Languages* 4, 13. doi:10.3390/languages4010013

Lindgren, R., and Johnson-Glenberg, M. (2013). Emboldened by Embodiment. *Educ. Res.* 42, 445–452. doi:10.3102/0013189x13511661

Lindgren, R., Tscholl, M., Wang, S., and Johnson, E. (2016). Enhancing Learning and Engagement through Embodied Interaction within a Mixed Reality Simulation. *Comput. Edu.* 95, 174–187. doi:10.1016/j.compedu.2016.01.001

Mathewson, J. H. (1999). Visual-spatial Thinking: An Aspect of Science Overlooked by Educators. *Sci. Ed.* 83, 33–54. doi:10.1002/(sici)1098-237x(199901)83:1<33:aid-sce2>3.0.co;2-z

Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., and Davis, T. J. (2014). Effectiveness of Virtual Reality-Based Instruction on Students' Learning Outcomes in K-12 and Higher Education: A Meta-Analysis. *Comput. Edu.* 70, 29–40. doi:10.1016/j.compedu.2013.07.033

Mou, W., and McNamara, T. P. (2002). Intrinsic Frames of Reference in Spatial Memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 162–170. doi:10.1037/0278-7393.28.1.162

Newcombe, N. S., and Shipley, T. F. (2015). "Thinking about Spatial Thinking: New Typology, New Assessments," in *Studying Visual and Spatial Reasoning for Design Creativity* (Springer), 179–192. doi:10.1007/978-94-017-9297-4_10

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

Ormand, C. J., Manduca, C., Shipley, T. F., Tikoff, B., Harwood, C. L., Atit, K., et al. (2014). Evaluating Geoscience Students' Spatial Thinking Skills in a Multi-Institutional Classroom Study. *J. Geosci. Edu.* 62, 146–154. doi:10.5408/13-027.1

Paas, F., and Sweller, J. (2012). An Evolutionary Upgrade of Cognitive Load Theory: Using the Human Motor System and Collaboration to Support the Learning of Complex Cognitive Tasks. *Educ. Psychol. Rev.* 24, 27–45. doi:10.1007/s10648-011-9179-2

Plummer, J. D., Bower, C. A., and Liben, L. S. (2016). The Role of Perspective Taking in How Children Connect Reference Frames when Explaining Astronomical Phenomena. *Int. J. Sci. Edu.* 38, 345–365. doi:10.1080/09500693.2016.1140921

Repetto, C., Serino, S., Macedonia, M., and Riva, G. (2016). Virtual Reality as an Embodied Tool to Enhance Episodic Memory in Elderly. *Front. Psychol.* 7, 1839. doi:10.3389/fpsyg.2016.01839

Ritzwoller, M. H., Barmin, M. P., Villaseñor, A., Levshin, A. L., and Engdahl, E. R. (2002). Pn and Sn Tomography across Eurasia to Improve Regional Seismic Event Locations. *Tectonophysics* 358, 39–55. doi:10.1016/s0040-1951(02)00416-x

Ruscella, J., and Obeid, M. F. (2021). "A Taxonomy for Immersive Experience Design," in 2021 7th International Conference of the Immersive Learning Research Network (iLRN) (IEEE), 1–5. doi:10.23919/ilrn52045.2021.9459328

Sanchez-Vives, M. V., and Slater, M. (2005). From Presence to Consciousness through Virtual Reality. *Nat. Rev. Neurosci.* 6, 332–339. doi:10.1038/nrn1651

Schreier, M. (2012). *Qualitative Content Analysis in Practice*. London: Sage publications.

Schubert, T. W. (2009). A New conception of Spatial Presence: Once Again, with Feeling. *Commun. Theor.* 19, 161–187. doi:10.1111/j.1468-2885.2009.01340.x

Schuemie, M. J., Van Der Straaten, P., Krijn, M., and Van Der Mast, C. A. P. G. (2001). Research on Presence in Virtual Reality: A Survey. *CyberPsychology Behav.* 4, 183–201. doi:10.1089/109493101300117884

Shapiro, L. (2014). *The Routledge Handbook of Embodied Cognition*. New York, NY: Routledge.

Shapiro, L. (2007). The Embodied Cognition Research Programme. *Philos. Compass* 2, 338–346. doi:10.1111/j.1747-9991.2007.00064.x

Skulmowski, A., and Rey, G. D. (2018). Embodied Learning: Introducing a Taxonomy Based on Bodily Engagement and Task Integration. *Cogn. Res.* 3, 6. doi:10.1186/s41235-018-0092-9

Slater, M. (2009). Place Illusion and Plausibility Can lead to Realistic Behaviour in Immersive Virtual Environments. *Phil. Trans. R. Soc. B* 364, 3549–3557. doi:10.1098/rstb.2009.0138

Slater, M., Spanlang, B., and Corominas, D. (2010). Simulating Virtual Environments within Virtual Environments as the Basis for a Psychophysics of Presence. *ACM Trans. Graph.* 29, 1–9. doi:10.1145/1778765.1778829

Slater, M., and Wilbur, S. (1997). A Framework for Immersive Virtual Environments (Five): Speculations on the Role of Presence in Virtual Environments. *Presence: Teleoperators & Virtual Environments* 6, 603–616. doi:10.1162/pres.1997.6.6.603

Smyrnaiou, Z., Sotiriou, M., Georgakopoulou, E., and Papadopoulou, O. (2016). Connecting Embodied Learning in Educational Practice to the Realisation of Science Educational Scenarios through Performing Arts. *Inspiring Sci. Edu.* 31, 31–38.

Southgate, E. (2020). "Conceptualising Embodiment through Virtual Reality for Education," in 2020 6th International Conference of the Immersive Learning Research Network (iLRN) (IEEE), 38–45. doi:10.23919/ilrn47897.2020.9155121

Stolz, S. A. (2015). Embodied Learning. *Educ. Philos. Theor.* 47, 474–487. doi:10.1080/00131857.2013.879694

Tarbuck, E. J., Lutgens, F. K., and Tasa, D. (1997). *Earth Science*. New Jersey: Prentice-Hall.

Venzke, E. (2013). Global Volcanism Program. *Volcanoes of the World* 4, 1.

Vorderer, P., Wirth, W., Gouveia, F. R., Biocca, F., Saari, T., Jäncke, L., et al. (2004). Mec Spatial Presence Questionnaire. Available at: http://www.ijk.hmt-hannover.de/presence (Accessed September 18, 2015).

Waller, D., and Greenauer, N. (2007). The Role of Body-Based Sensory Information in the Acquisition of Enduring Spatial Representations. *Psychol. Res.* 71, 322–332. doi:10.1007/s00426-006-0087-x

Weise, M., Zender, R., and Lucke, U. (2019). "A Comprehensive Classification of 3d Selection and Manipulation Techniques," in Proceedings of Mensch und Computer (IEEE), 321–332. doi:10.1145/3340764.3340777

Wilson, M. (2002). Six Views of Embodied Cognition. *Psychon. Bull. Rev.* 9, 625–636. doi:10.3758/BF03196322

Wirth, W., Hartmann, T., Böcking, S., Vorderer, P., Klimmt, C., Schramm, H., et al. (2007). A Process Model of the Formation of Spatial Presence Experiences. *Media Psychol.* 9, 493–525. doi:10.1080/15213260701283079

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

Yeonhee, C. (2018). *How Spatial Presence in VR Affects Memory Retention and Motivation on Second Language Learning: A Comparison of Desktop and Immersive VR-Based Learning*. Syracuse, NY: Syracuse University. Ph.D. thesis.

Yuan, L., Kong, F., Luo, Y., Zeng, S., Lan, J., and You, X. (2019). Gender Differences in Large-Scale and Small-Scale Spatial Ability: A Systematic Review Based on Behavioral and Neuroimaging Research. *Front. Behav. Neurosci.* 13, 128. doi:10.3389/fnbeh.2019.00128

Zielasko, D., Horn, S., Freitag, S., Weyers, B., and Kuhlen, T. W. (2016). "Evaluation of Hands-free Hmd-Based Navigation Techniques for Immersive Data Analysis," in 2016 IEEE Symposium on 3D User Interfaces (3DUI) (IEEE), 113–119. doi:10.1109/3dui.2016.7460040

Zielasko, D., and Riecke, B. E. (2021). To Sit or Not to Sit in VR: Analyzing Influences and (Dis)Advantages of Posture and Embodied Interaction. *Computers* 10, 73. doi:10.3390/computers10060073

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Virtual Reality

ORIGINAL RESEARCH
published: 18 February 2022
doi: 10.3389/frvir.2022.742509

The Interplay Between Presence and Learning

Carli Ochs[1] *and Andreas Sonderegger*[1,2]*

[1]*Department of Psychology, University of Fribourg, Fribourg, Switzerland,* [2]*Business School, Institute for New Work, Bern University of Applied Sciences, Bern, Switzerland*

The highly immersive Virtual reality (VR) headset is gaining popularity in multiple application domains. In the context of learning, it has been proposed to be beneficial by increasing presence and attention in noisy and distracting environments, both factors that are considered important for learning. Despite intensified research efforts in recent years, empirical knowledge of experimental research addressing the link between presence and learning in specific environmental contexts is still rather scarce. In this study following an experimental mixed-method approach, the link between presence and memorization as a particular form of learning is addressed by comparing memorization with a highly immersive VR headset to a less immersive system (desktop screen) in noisy and calm learning environments. Using a 2 (learning location) x 2 (learning device) between-subjects design, 63 participants interacted with one of the two devices in either of the two environments. As expected, VR headset users reported higher presence levels. While participants subjectively evaluated the VR headset as a better device for learning, the memorization test scores were higher for desktop screen users in both calm and noisy environments. Learning location did not show significant effects. Attention distraction and context-dependent learning are discussed with regard to the unexpected results, while implications for practice and future research are discussed.

Keywords: presence, memorization, learning, virtual reality, immersion, environment, noise

OPEN ACCESS

Edited by:
Richard Skarbez,
La Trobe University, Australia

Reviewed by:
Aleshia Hayes,
University of North Texas, United States
George Papagiannakis,
University of Crete, Greece

***Correspondence:**
Andreas Sonderegger
andreas.sonderegger@bfh.ch

Specialty section:
This article was submitted to Virtual Reality and Human Behaviour, a section of the journal Frontiers in Virtual Reality

Received: *16 July 2021*
Accepted: *14 January 2022*
Published: *18 February 2022*

Citation:
Ochs C and Sonderegger A (2022) The Interplay Between Presence and Learning. Front. Virtual Real. 3:742509. doi: 10.3389/frvir.2022.742509

1 INTRODUCTION

In the past few years, a new trend has emerged in line with the commercialization of a new device, allowing for complete immersion in a virtual world (Lombard, 2016; Skarredghost, 2017). Virtual reality (VR) technology is opening up a plethora of new opportunities. It is being used in exposure therapy (Price and Anderson, 2007; Parsons and Rizzo, 2008), treatment of addiction, and post-traumatic stress disorder (Beck et al., 2007; Baños et al., 2011), military training (Rizzo et al., 2011), medical training (Piedra et al., 2016; Rahm et al., 2016), museums (Sylaiou et al., 2010), tourism (Huang et al., 2016; Potdevin et al., 2021), and other domains (Merchant et al., 2014). Another promising application domain for VR technology is the field of education and training (Pan et al., 2006; Hayes et al., 2013; Stevens and Kincaid, 2015; Cryer et al., 2019). In this context, the question arises as to whether the use of VR technology has a beneficial influence on learning outcomes. Although this question has been addressed in previous research, the number of methodologically sound studies assessing learning outcomes objectively are rather scarce (Radianti et al., 2020). In addition, still very little is known regarding the underlying processes influencing the link between VR use and learning. In this regard, experience of presence is often referred to as explanation for a positive link between VR use and learning. However, the empirical findings are inconclusive, with studies indicating positive effects of presence on student learning while others reported the opposite effect (Jensen and Konradsen, 2018; Radianti et al., 2020). Several factors have been put forward that might have an influence on the interplay between presence and learning such as the experience of cybersickness, type of content to be learned and the environment or context in which such learning takes place (Jensen and Konradsen, 2018; Radianti et al., 2020). In this respect, this study sets a focus on the interplay between the experience of presence as a consequence of VR use, learning environment, cybersickness and learning outcomes in the context of cognitive skills acquisition.

1.1 Immersion, Presence and Virtual Reality

VR is a computer technology composed of hardware and software that simulates a physical presence of a user in an environment that is virtually created (Biocca and Delaney, 1995; Radianti et al., 2020). While different definitions of VR have been put forward, a classification often referred to differentiates between low and high immersive VR (Biocca and Delaney, 1995; Lee and Wong, 2014). Low immersive VR relates to technological devices such as a desktop computer screen while high immersion VR is most commonly associated with a head-mounted system (i.e., VR headset).

VR systems will meet a user's sensory expectations to different degrees (Alhalabi, 2016). The more the system meets these expectations, the more a person will be able to forget that she or he is using a VR system and will be able to appreciate its content (Burdea et al., 1996; Slater and Wilbur, 1997; Bowman and McMahan, 2007). The quality of the experience in virtual environments is generally referred to as presence or immersion. Although slightly different in terms of scope and definition, the two concepts are closely linked: if immersion is high, presence will be high and vice versa (Bowman and McMahan, 2007).

There is some disagreement about the definition of immersion (Slater, 2003; Mütterlein, 2018). Some researchers consider immersion as a psychological state and define it as the feeling of being absorbed in the virtual world (Freina and Canessa, 2015; Mütterlein, 2018). Other researchers, us included, define immersion as an objective measure that reflects the level of sensory fidelity a VR system affords (Slater and Wilbur, 1997; Bowman and McMahan, 2007). It represents an objective indicator which depends on the hardware and/or software of the VR environment (e.g., the display size and resolution, the refresh rate, the realism of lighting, the type of position tracker, the quality of the visual, audio, and haptic feedback Burdea et al. (1996)).

Presence, on the other hand, is the subjective psychological response to a VR experience, reflecting the extent to which users actually feel part of the (virtual) world they are experiencing (Slater and Wilbur, 1997; Bowman and McMahan, 2007). Presence is defined as the successful experience of being somewhere else (Heeter, 1992; Butler, 1997). In other words, the level of immersion with a specific VR system should be the same for every user as it is an objective measure that depends on the software and hardware of the device. In contrast, the individual experience of presence is very subjective and depends heavily on specific individual appraisal processes (Bowman and McMahan, 2007).

Several studies have indicated that presence may have a considerable influence on emotional reactions (Riva et al., 2007), motivation (Vogel et al., 2006), pain management (Sharar et al., 2008), exposure therapy (Parsons and Rizzo, 2008), and training (Stevens and Kincaid, 2015). Several authors also suggested that presence might be linked to learning. Psotka (1995) and Makransky et al. (2019) suggest that the feeling of presence might increase students' motivation to learn which leads to better learning. Witmer and Singer (1998) argued that the positive link between presence and learning can be explained based on the reduction of distraction and increasing focus. In addition, they suggest that VR is beneficial for learning due to the increased level of immersion it offers (as compared to a desktop computer).

1.2 Presence and Learning

Defined as "the relatively permanent change in a person's knowledge or behaviour due to experience or practice" (Mayer, 1982), learning can be categorized into various levels of difficulty, with memorization (e.g., being able to recall facts and concepts) being the simplest and creating (e.g., producing new or original work) the most complex level (Anderson et al., 2001). In addition, various types of learning can be differentiated. In this regard, the distinction between the acquisition of cognitive, psychomotor and affective skills has been suggested (Jensen and Konradsen, 2018).

Potential positive benefits of presence for learning were already discussed before highly immersive devices were available on the consumer market (Psotka, 1995; Salzman et al., 1995; Mikropoulos and Natsis, 2011). Few studies however addressed this link in low immersion environments. One study reported that presence felt by participants using a desktop computer did have an effect on learning (Ai-Lim Lee et al., 2010). Regarding highly immersive VR, quite a few studies have been conducted in the context of learning. Most of these studies, however, were addressing issues of usability of educational software, in which neither presence nor learning was measured (Schofield, 2012; Hupont et al., 2015; Dolezal et al., 2017; dela Cruz and Mendoza, 2018; Veronez et al., 2018; Radianti et al., 2020). While the usability of a virtual learning environment is necessary for both successful learning and presence, it does not guarantee a sense of presence or successful learning.

To our knowledge, very few studies have assessed both presence and learning in low and highly immersive environments. Some studies only measured presence and others only measured learning. Hupont et al. (2015) for example showed that presence was higher in a highly immersive VR educational environment (e.g., head-mounted headset) than in a low immersive desktop computer environment without assessing learning. In a similar vein a different study measured presence, immersion and self-efficacy in a low and highly immersive VR environment and reported that while presence and immersion was significantly higher, self-efficacy did not significantly vary between devices (Shu et al., 2019). The consequences on learning however have not been assessed in that work.

In other studies, learning was assessed in different immersive environments, but not presence. Alhalabi (2016) for example, compared the learning performance of students using four different devices (e.g., high immersive VR headset with head tracking, high immersive corner cave system, low immersive VR headset without head tracking, and low immersive computer screens). While students performed best in the most immersive system (which made the authors conclude that immersion was positively influencing learning), presence was not measured. In a similar study comparing learning with three devices (i.e., desktop computer, cave system and VR headset), VR users performed the worst. While the paper suggest that the novelty of immersion might have led to lower score, presence was also not measured. In another study (Kozhevnikov et al., 2013), participants learned relative motion concepts either on a desktop computer or with a VR headset. VR headset users had a higher learning score, which was interpreted by the authors to be a consequence—among others—of the higher immersion of VR devices, which however was not assessed. Learning was also shown to be higher for VR users in a study comparing learning on the topic of vocational training either with a google carboard VR headset or in a conventional classroom setting (Ray and Deb, 2016). Another study using google carboard VR headset, users were reported to specifically do better on questions regarding spatial awareness (Rasheed et al., 2015). Lastly since the start of the COVID-19 pandemic some studies have also looked into the possibility of teaching medical skills thanks to VR, with mixed results in terms of learning (Lohre et al., 2020; Birrenbach et al., 2021; Clarke, 2021; Pears and Konstantinidis, 2021). However, in none of these studies presence was measured. Therefore, is not clear what processes led to differences in learning performance comparing the different learning environments.

Astonishingly, two studies that did asses presence and learning in low and highly immersive learning environments (Moreno and Mayer, 2004; Makransky et al., 2019) showed that while presence was higher in the highly immersive VR device, learning scores were lower. The authors assumed that these results were due to a higher cognitive workload that was linked to the novelty of the VR headset. In a similar vein, findings of other studies indicated that while presence was higher in a more immersive environment (cave system vs interactive workbench and PC vs head mounted display), learning (referred to as memorization) was not affected by immersiveness of device, with no significant link between presence and memorization (Sutcliffe et al., 2005; Buttussi and Chittaro, 2018). In contrast to this, Cadet and Chainay (2020) reported that both presence as well as memorization were significantly higher for participants using the more immersive device (VR-headset vs computer scren; Cadet and Chainay, 2020). In a similar vein, Stevens and Kincaid (2015) reported a moderate positive correlation between the degree of presence and learning performance (Stevens and Kincaid, 2015). Interestingly another more recent study found that while device (screen vs VR headset) showed no effect on learning performance, there was a link between presence and performance, with participants subjectively reporting higher levels of presence also obtaining higher scores on a performance test (Grassini et al., 2020).

Hence, it can be summarized that the state of knowledge regarding the relation between presence and learning is inconclusive. In theory, presence has been argued to be beneficial for learning, suggesting VR to be a useful technology in this context. Studies indicating a positive relation between VR and learning hence often refer to increased presence as explanation for the effect, without explicitly assessing it. In contrast to this, presence did not show a positive relation with learning in several studies assessing both, learning and presence empirically. Therefore, additional research addressing the link between presence and learning is needed, taking into consideration other potential influencing factors such as the learning environment or negative consequences of technology use such as cybersickness (Psotka, 1995; Witmer and Singer, 1998; Salzman et al., 1999; Alhalabi, 2016; Jensen and Konradsen, 2018).

1.3 Additional Factors Influencing Learning in Virtual Reality

1.3.1 Learning Environment

With regard to training and education, an advantage to VR environments is the control they provide over the contextual learning environment. This is because contextual elements such as temperature (Hutchinson, 2003), noise (Haines et al., 2001; Cassidy and MacDonald, 2007), lighting (Haines et al., 2001), air quality (Wyon, 2004), furniture setting, and visual cues (Davis, 1984) have been shown to have significant effects on learning.

Although it is not yet possible to control all elements of the learning environment in a VR application (e.g., temperature or air quality are difficult to influence), aspects such as lighting, colour, noise, and visual cues can be designed in order to create an ideal environment for learning. Especially with regard to disturbances due to visual cues, VR head mounted systems might come with an extra benefit for learning. In highly-immersive VR learning environments, students wear a headset, which forces them to focus on the content provided in the VR-environment. Since they cannot see anything else, it can be expected that they are less prone to distractions compared to a classical learning environment (Salzman et al., 1999; Bowman and McMahan, 2007). Because any form of distraction in the environment (e.g., visual, physical or auditory) has been shown to create short interruptions in attention that impinge on individual learning outcomes (Altmann et al., 2014), this isolation from the exterior world in a VR environment can be a decisive advantage over conventional learning environments. This is particularly the case taking into consideration the fact that nowadays, most people receive pop up messages on their phone or computer that distract them and entice them to multitask, which has been shown to negatively influence learning (Spira and Feintuch, 2005; Winter et al., 2010).

In current VR environments, the user is fully immersed in the environment he or she is learning in (Salzman et al., 1995; Fried, 2008; Winter et al., 2010; Fisher et al., 2014). Although the isolation from external disturbing cues seems to be an obvious advantage of highly-immersive VR systems, very limited evidence of empirical research is available addressing this question by comparing high and low immersive technology while varying disturbance potential of the learning environment.

1.3.2 Cybersickness

One potential influencing factor regarding the link between increased presence as consequence of a highly immersive virtual environments and learning is cybersickness. Cybersickness, also referred to as virtual reality induced symptoms and effects (VRISE, e.g., Nichols et al., 1997) or negative effects (e.g., Lessiter et al., 2001), is a subtype of motion sickness that is used to describe symptoms of discomfort and illness caused by VR (Mazloumi Gavgani et al., 2018). Possible manifestations of cybersickness are nausea, vomiting, fainting, dizziness, a sensation of spinning, sweating, feeling hot, tiredness, annoyance, and blurred vision (e.g., LaViola, 2000).

The sensory mismatch theory has been put forward in order to explain both cybersickness and motion sickness (Mazloumi Gavgani et al., 2018). The theory posits that motion sickness occurs when conflicting signals are received from the spatial orientation senses (e.g., the vestibular system, the eyes, and the non-vestibular proprioceptors; Bles et al., 1998). Cybersickness is most likely caused by a mismatch between visual stimuli and the appropriate vestibular or proprioceptive feedback (Mazloumi Gavgani et al., 2018).

Some factors are known to increase the likelihood of experiencing cybersickness (Mousavia et al., 2013). Some of these are related to technical issues such as flickers, lags, and position tracking errors. While others are related to individual difference such as gender, age, and illness (e.g., Sharples et al., 2008; Davis et al., 2014). Children from the age of 2–12, women and individuals experiencing some form of illness are more likely to be affected by cybersickness. Cybersickness has been shown to negatively influence presence (Lessiter et al., 2001; Polcar and Horejsi, 2015; Weech et al., 2019), and might hence play an important role in the context of VR and learning. However a previous study comparing two different VR headset devices has shown that cybersickness had no effect on learning (Moro et al., 2017).

1.4 The Present Study

VR allows for the simulation of real-life scenarios and create immersive and captivating experiences. This is why this technical environment has often been considered to be the future of teaching and training (Abulrub et al., 2011; Fisher et al., 2014). While there seems to be evidence for an increased use of VR-technology in teaching and education (e.g., Lessiter et al., 2001; Carrozzino and Bergamasco, 2010; Radianti et al., 2020) it is still not clear how the use of VR is related to learning performance and to what extent presence as underlying mechanism plays a role in this relation. In addition, only very little is known so far regarding the usefulness of VR in different learning environments. While learners in a classical learning environment using a desktop computer may be distracted by visual and auditory cues of their environment, VR isolates learners from contextual distractors. As distraction has been negatively linked with learning performance in previous research, VR users are expected to show better performances in a learning task compared to computer users.

In order to address these open questions, an experiment was conducted in which participants either used VR or a desktop computer screen for a learning task. Presence and cybersickness were assessed in order to better understand the underlying processes linked with the use of VR for learning. In order to reduce a potential novelty effect, participants were given time to familiarize themselves with the virtual environment. In addition, the level of contextual distraction was manipulated experimentally. Half of the participants completed the learning task in a calm and quiet room with no visual and auditory distractions. The other half learned in the entry hall of the university cafeteria, a noisy space with a lot of activity going on. The experiment being quite short we conceptualized learning as the simplest category of Bloom's taxonomy of educational objectives: participants were only expected to recall facts and basic concepts (Bloom et al., 1984). We therefore tested their learning performance with a memory test on information participants had seen in the virtual environment. It was expected that learning performance would be highest for VR users in the lab, followed by VR-users in the cafeteria and computer users in the lab, while computer users in the cafeteria were expected to show the weakest learning performance.

2 METHOD

2.1 Participants

Sixty-five participants aged 18 to 26 (M = 21.27, SD = 1.64) took part in the study (N_{female} = 50). In order to obtain a homogeneous sample, only university students from the University of Fribourg were recruited for the study. Participants could receive course credits in exchange for participation. Data of two participants were removed; one, due to technical problems during data collection, the other because the language of the experiment was not the mother tongue of the participant (which was a prerequisite for participation in the study). An additional exclusion criterion, participants had to meet was not to be dyslexic.

We also controlled participants' level of exposure to media devices (TV, Virtual reality headset, video games) that could influence their experience with VR. Of the 63 participants, 19 had already used a VR headset. From these, none had used one more than four times. Participants reported their level of expertise with a VR headset on a five-point Likert scale (1 = novice and 5 = expert) with a mean of 3.27 (SD = 0.77). On average, participants, watched 6.27 h of TV a week (SD = 0.679). Only 4.8% of the participants reported playing video games daily while 57.1% reported never playing any video games.

2.2 Experimental Design

The experiment followed a 2 by 2 between-subjects design, with location and immersion as independent factors. The participants either did the experiment in the university cafeteria (N = 32) or a

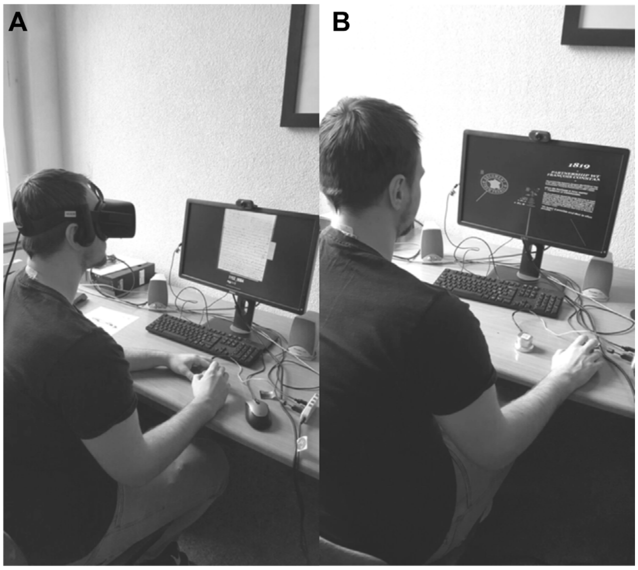

FIGURE 1 | Example participant (actor) in **(A)** the VR headset condition, **(B)** the desktop computer condition.

lab (N = 31). They either completed the memorization tasks with a desktop computer screen (N = 30) or with a VR headset (N = 33). **Figure 1A** shows an actor representing a participant working with a VR headset while in **Figure 1B**, the same actor can be seen working in the desktop computer condition.

2.3 Material

2.3.1 Hardware

A desktop system (Apple Mac Pro A1481) was used with either an immersive VR system (Oculus Rift CV1) or a 17 computer screen (DELL monitor E916H, resolution 1366 x 768 pixels). Participants using the VR only used a controller (Griffin™ PowerMate Programmable Multimedia Controller). It allowed them to select documents, advance on the timeline and zoom into documents. Participants using the virtual environment on a desktop computer screen also had to use a mouse (Microsoft L2 Comfort Mouse 4500) to change visual fields. **Figure 2** depicts both the controller and mouse.

In addition, loudspeakers were used to play the pre-recorded instructions for the memorization tasks. A decibel meter (Voltcraft SL-100) was used to measure the sound level in all experimental conditions.

2.3.2 The Virtual Environment

The environment the participants used was a virtual archive for historical documents of a traditional watch-manufacturer (**Figure 3**). The documents appeared on the sides of a timeline in chronological order. The participants could travel the timeline in both directions and select documents. The documents and the timeline were represented in an abstract black three-dimensional space.

The documents were organized by country of origin of the document. There was a main timeline with documents originating from Switzerland. From this timeline, it was possible to get to timelines from other countries (e.g., French timeline contained only documents that originated from France). The content of the documents were letters, pictures, contracts, flyers, and newspaper articles.

FIGURE 2 | left is mouse, right is controller.

2.4 Measures

2.4.1 Virtual Reality Experience and Media Usage

We asked a few questions in order to control for eventual differences of VR effects. We asked participants to estimate how much television they watched ("Please estimate your daily television consumption (hours/minutes)."). Participants were as well asked to rate their frequency of video game playing ("How frequently do you play video games?") on a Likert scale 1 (never) to 5 (every day). They were also asked to rate on a Likert scale ranging from 1 (none) to 5 (expert), their experience with computers ("Please rate your level of experience with computers") as well as their knowledge on VR ("How would you rate your level of knowledge on VR? (e.g., how it functions)."). Lastly, participants were asked if they had ever used a VR headset before ("Have you already used a VR headset before?") and if so, how often ("If yes, how often have you used it?")

2.4.2 Presence

To measure spatial presence we used an adapted version of the ITC-Sense of Presence Inventory (ITC-SOPI) (Lessiter et al., 2001). In contrast to most other presence questionnaires, the ITC-SOPI was created to measure presence independently from the type of device and environment that is used (TV, computer, VR headset, etc.). The original questionnaire is composed of 44 items that can be divided into four factors: spatial presence, ecological validity, engagement and negative effects (i.e., cybersickness). The questionnaire uses a five-point Likert scale (ranging from completely disagree to strongly agree). In this study we only used negative effects (i.e., cybersickness) and spatial presence which together encompassed 25 items.

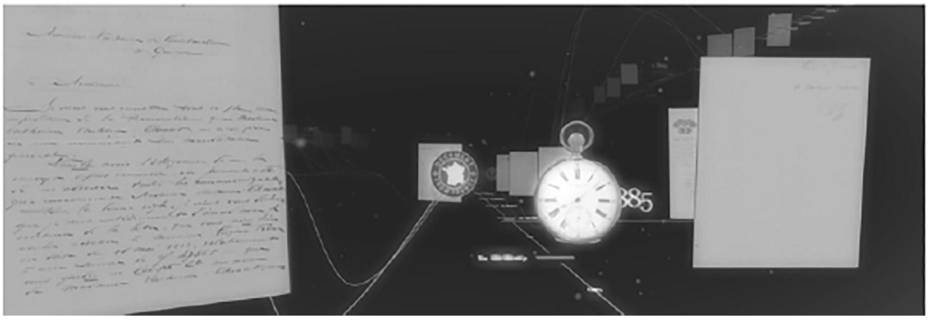

FIGURE 3 | VR archives of a watch-manufacturer used for the memorization task.

Spatial Presence

The spatial presence dimension refers to the participant's sensation of "being there". Thirteen items of the spatial presence scale were removed because they were irrelevant for this study. Example of reasons for exclusion are referring to non-existing characters (e.g., "I had the sensation that characters were aware of me.") or non-existing smells (e.g., I could almost smell different features of the displayed environment). The items that were selected for spatial presence can be found in **Table 1**. The deletion of the inappropriate items left a scale of five items. The items could be answered on a Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The reliability for this shortened Spatial Presence scale (5 Items) was 0.76 (Cronbach's alpha).

Cybersickness

The subscale "negative effects" of the ITC-SOPI is referred to in this study as cybersickness and measures adverse physiological reactions while using a VR headset (e.g., dizziness, headache). The scale is composed of six items. Each item could be answered on a Likert scale ranging from (strongly disagree) to (strongly agree). The reliability for the for cybersickness scale was 0.9 (Cronbach's alpha).

2.4.3 Memorization Test

A memorization test was developed on the basis of the information presented in the digitalized archive. The tasks the participants completed (cf. procedure section) ensured that participants read all the respective text in the archive to be able to answer the test. The test was composed of 21 multiple choice questions with an option of four answers, with only one answer being correct. Six questions asked about the date and locations of the documents (i.e., when and where something happened). Five questions were asked about the number of documents represented in a specific area in the VR environment. Four questions were about the type of document which was viewed. Finally, seven questions were asked about the content of the diverse documents (i.e., what was said). The average item difficulty was $M = 0.58$, $SD = 0.2$, with the most difficult item having a difficulty level of 0.3 and the easiest being at 1.

TABLE 1 | Items of the short spatial presence scale.

1. I felt I could interact with the displayed environment
2. I felt I was visiting the places in the displayed environment
3. I felt I wasn't just watching something
4. I felt surrounded by the displayed environment
5. I felt I could have reached out and touched things (in the displayed environment)
6. I felt that all my senses were stimulated at the same time

2.4.4 Disturbance due to Location

A four-item scale was created to measure the possible level of experienced disturbance by the participants depending on the location. Participants could answer on a Likert scale ranging from 1 (not at all) to 5 (a lot). Two items asked questions regarding the sound level ('was there any noise during the experiment?' "Were you disturbed by the noise?") and two other items were linked to the presence of others ("were you disturbed by the presence of others during the experiment?" "Did you feel uncomfortable during the experiment?")

2.4.5 Comparative Questions on the Two Devices

At the end of the experiment, after participants had interacted with the other device (e.g., VR for those who learned with the computer, cf. procedure section), a final questionnaire was administered containing five questions comparing the two devices (e.g., "Which of the two devices was easier to use?"). The five questions can be found in **Table 3**. A semantic differential scale ranging from one (computer screen) to five (VR) was used. In addition, after each of the five questions, participants were asked to justify their rating by entering a short comment as free text.

2.5 Procedure

Participants were either received in the lab or the cafeteria. They were first asked to sign the consent form and were informed about the procedure of the experiment. They then filled out a questionnaire containing the demographic and control questions. After this, they were instructed in detail on how to use the device they were going to use (desktop computer screen or virtual reality headset). This was done so that users wouldn't encounter usability issues during the experiment. In a warm-up phase, participants were asked to familiarize themselves with the device for 3 minutes. We did this to avoid concerns put forward by Moreno and Mayer (2004) and Makransky et al. (2019) about participants not performing well in VR due to excitement and distraction in the new and unfamiliar virtual environment, to which we refer to as the “wow-effect”.

After the warm-up phase, participants received instructions for the following tasks. They were informed that questions would be asked later about their task. In addition, they were reminded to stay concentrated throughout the whole experiment. Then, the facilitator left the laboratory or the cafeteria and went to a neighbouring room, from where she could see and hear what the participant was doing by means of a screen-mirroring application (TeamViewer). Participants received all information regarding the different tasks via pre-recorded audio-instructions. If needed, participants could ask to have the task instructions repeated by saying “repeat”.

After task completion, participants were asked to answer a series of questionnaires including the spatial presence and negative effects (i.e., cybersickness) dimensions from the ITC-SOPI, the test on the tasks, the questions on annoyance and finally the ones on satisfaction.

After this, participants changed the device (e.g., computer screen for participants who completed the memorization tasks in the VR environment) and interacted with the alternative device for 3 minutes. They then replied to the questions comparing the two devices.

2.6 Data Analysis and Manipulation Check

Data was analysed using a two-factorial analysis of variance (ANOVA). For some variables, the Kolmogorov Smirnoff test indicated a non-normal distribution and the assumption of homogeneity of variance was not met (short presence 2, disturbance, satisfaction and learning). Since the QQ plots showed for most of these variables a normal distribution and since the absolute size of the skewness and kurtosis relative to their standard error was always below 2, it was assumed that the ANOVA can be considered robust (Khan and Rayner, 2003).

Results of the manipulation check regarding location can be found in **Table 2**. Manipulation check indicated that the physical sound measured (in decibels) was louder in the cafeteria (M = 55,59, SD = 11.31) than in the lab (M = 30,14, SD = 10,48) indicating a significant manipulation of noise F (1,58) = 81.51, $p < 0.05$, $\eta 2$ = 0.58. There was also a positive correlation between the level of sound and the perceived level of sound $r = 0.60$, $N = 60$, $p < 0.05$, as well as between the level of sound and the level of disturbance through sound $r = 0.51$, $N = 60$, $p < 0.05$.

The level of disturbance was not affected by the device used by the participant F (1, 53) = 2,1 $p > 0.05$, $\eta 2$ = 0.04. Participants using VR headset (M = 1.9, SD = 1.23) reported a similar level of disturbance to desktop computer screen users (M = 1.51, SD = 0.74).

3 RESULTS

3.1 Memorization Test

Results in the memorization test showed, in contrast to our expectations, that participants in the desktop computer screen condition performed better (M = 12.57, SD = 1.87) than participants in the VR headset condition (M = 10.73, SD = 2.71; F (1, 58) = 9.51 $p < 0.05$, $\eta^2 = 0.14$). In order to control for the influence of cybersickness, the same analysis was run with cybersickness as a co-variate. Results showed that desktop users still showed a higher memorization score than the VR headset users, but the size of this effect was considerably smaller F (1, 57) = 5.38 $p < 0.05$, $\eta^2 = 0.09$.

Location had no significant effect on learning F (1, 59) = 0.21, $p > 0.05$, $\eta^2 = 0.01$. Participants in the cafeteria (M = 11.5, SD = 2.99) showed similar learning scores compared to the ones in the lab (M = 11.71, SD = 1.92). There was no significant interaction of device and location on the memorization score F (1, 59) = 0.79, $p > 0.05$, $\eta^2 = 0.01$.

3.2 Presence

3.2.1 Spatial Presence

Results showed that there was a significant main effect of device indicating higher spatial presence scores in the VR condition (M = 3.54, SD = 0.6) compared to the computer condition (M = 3.16, SD = 0.72; F (1, 59) = 5.76, $p < 0.05$, $\eta^2 = 0.09$). Location did not show a significant main effect F (1, 59) = 2.05, $p > 0.05$, $\eta^2 = 0.034$. There was also no significant interaction between the two factors F (1, 59) = 0.003 $p > 0.05$, $\eta^2 = 0.00$.

3.2.2 Cybersickness

Cybersickness was higher in VR headset users (M = 3, SD = 1.16) compared to desktop users (M = 1.93, SD = 0.76; F (1, 59) = 18.37 $p < 0.05$, $\eta^2 = 0.87$). There was no significant difference between participants in the cafeteria (M = 2.52, SD = 1.2) and in the lab condition (M = 2.46, SD = 1.04; F (1, 59) = 0.11, $p > 0.05$, $\eta^2 = 0.00$). There was also no significant interaction between the two factors F (1, 59) = 3.37, $p > 0.05$, $\eta^2 = 0.05$. An analysis of the above-reported triple interaction on spatial presence using cybersickness as co-variate did not result in any changes with regard to the presented results.

3.3 Correlation of Main Variables

Analysis of correlations between our main variables (memorization test, spatial presence, and cybersickness) revealed a significant negative link between cybersickness and the score on the memorization test (cf. **Table 3**). Spatial presence, however, was not significantly linked to memorization or cybersickness. As can be seen in **Table 3**, when running the analysis for each of the two immersion conditions separately, there was no significant link between the different variables.

TABLE 2 | Results for disturbance items.

	Location		F(1, 61)	p	η2
	Lab	Cafeteria			
	M(SD)	M(SD)			
Perceived noise level	1.38 (0.57)	3.23 (1.28)	43.05	0.000	0.45
Disturbance due to noise level	1 (0.00)	2.61 (1.31)	36.32	0.000	0.41
Disturbance by others	1 (0.00)	2.42 (1.32)	28.02	0.000	0.35
Feeling uncomfortable	1 (0.00)	1.74 (1)	13.18	0.001	0.19
Total mean	1.09 (0.14)	2.5 (1.10)	38.34	0.000	0.42

TABLE 3 | Correlations of the main variables for participants in VR and computer condition.

Measures		Spatial presence	Cybersickness
Memorization score	VR (*N* = 33)	0.190	–0.093
	Computer (*N* = 30)	0.107	–0.200
	Total (*N* = 63)	0.043	–0.278*
Spatial presence	VR (*N* = 33)	—	–0.115
	Computer (*N* = 30)	—	0.061
	Total (*N* = 63)	—	0.082

*$p < 0.05$.

3.4 Comparative Evaluation of the Two Devices

Participants interacted with both devices (cf. method section). This made it possible to ask participants, at the end of the study, to reply to items addressing their beliefs on the benefits and disadvantages of these two devices for learning and studying. In addition to the comparative semantic differentials, written comments were screened and categorized by one of the authors. Two coders then coded participants' answers into the categories defined by the author. In case of disaccord, the statement was discussed by the two raters in order to obtain a consistent assessment.

As can be seen in **Table 4**, a majority of participants preferred the VR headset to the desktop computer. Reasons for preferring VR for learning given by participants were being more immersed with VR (mentioned by 46.03% of participants), feeling more implicated (14.28%), finding VR more interesting (12.69%) and interactive (11.11%). In terms of device preferences for work, the results are more nuanced. Participants that reported preferring VR for work explained that they felt more immersed with VR (17.46%), that VR was easier to use (11.11%) and that they would be more concentrated with VR (11.11%). Participants that though they would work better with the computer listed the absence of cybersickness symptoms (12.7%) and their previous knowledge with computers (12.7%) as reasons. Cybersickness symptoms were also listed as an explanation for believing that they would be more focused on a desktop computer (12.69%). Additionally, some participants expressed the belief that VR was for fun and not for work (15.87%). Participants who believe that they would be most concentrated with the VR headset listed immersion (31.74%) and being physically cut off from the outside world (23.8%) as reasons.

In terms of ease of use (cf. **Table 4**), participants preferring VR explained that VR felt more intuitive and that it was easier to orient oneself. Several participants (15.85%) mentioned that the desktop computer was easier to use because of previous experience.

There was no clear preference with regard to the question about the device's usefulness for information retention (cf. **Table 4**). However, only VR users gave explanations about their reasoning, mentioning higher immersion (17.46%), implication (17.46%) and focus (19.05%) as reasons for why they prefer VR over desktop computers regarding information retention.

4 DISCUSSION

The primary hypothesis of this study was that students using a VR headset would feel more present and hence learn better compared to participants using a desktop computer. The difference between the groups was expected to be even stronger if the participants took part in the experiment in the noisy and distracting cafeteria instead of the calm and quite lab.

As expected, students reported higher values of spatial presence when using a VR headset compared to the desktop screen users. Contrary to our expectations, however, this had no beneficiary influence on participants' memorization performance, since data of the memorization test showed that participants learned better when using a computer screen. Interestingly, this also contradicts participants' subjective evaluation of the usefulness of the two technological devices for learning; 65% of the participants reported that they expected to learn better using a VR headset compared to the desktop computer. These findings are is in stark contrast to the

TABLE 4 | Device preferences.

	VR	Same			Computer screen
	(1)	(2)	(3)	(4)	(5)
What device did you prefer?	54%	17.5%	6.3%	15.9%	6.3%
With which device do you think you would work better with?	22.2%	20.6%	17.5%	19%	20.6%
With which device do you think you would work the most concentration?	25.4%	27%	19%	17.5%	11.1%
With which device did you have the best ease of use?	20.6%	20.6%	31.7%	19%	7.9%
Which device would allow you to retain more information and why?	3.2%	12.7%	19%	12.7%	3.2%

widely accepted assumption that presence positively influences learning (Psotka, 1995; Salzman et al., 1995; Mikropoulos and Natsis, 2011). Interestingly, this is in line with results of other studies that have addressed presence and learning in highly immersive environments (Moreno and Mayer, 2004; Makransky et al., 2019). Several potential explanations can be put forward in this regard.

A first potential explanation for this unexpected result might be found in the data on cybersickness and disturbances. VR headset users reported increased disturbance levels (e.g., felt uncomfortable) and experienced cybersickness (e.g., dizziness, feeling sick, headaches). Also, in their qualitative feedback, participants mentioned cybersickness as main reason for preferring the desktop computer screen to the VR headset for working and memorization (cf. **Table 4**). However, presence seems to be affected by cybersickness only to a very limited extent, as the correlation between these two variables was rather small.

However, factors such as lag (delay between user action and system response), flicker, calibration (interpupillary adaptation), and general ergonomics (e.g., heavy and poor fitting headsets) have also been shown to have a considerable influence on the prevalence of negative consequences of VR use (e.g, McCauley and Sharkey, 1992; LaViola, 2000; Sharples et al., 2008). Previous research has shown that the occurrence of cybersickness was linked with lower attitudes towards the technology, lower learning outcomes and lower presence (Polcar and Horejsi, 2015; Weech et al., 2019). With regard to learning and VR, future research needs to address the question of how the improvement of hardware and software can help to reduce the occurrence of negative consequences in VR use, as this aspect seems to have an important influence on learning performance. Although such negative effects seem to impinge on learning with the VR headset (cf. the negative correlation between the two measures in this study), analysis of co-variance showed that when controlling for cybersickness, participants still perform better in the desktop computer screen condition—however with a reduced effect size. Therefore, it can be assumed that cybersickness linked with the use of the VR headset is not the only explanation for the reported differences in learning performance.

A second assumption explaining this unexpected effect might be found in differences in previous experience of the participants with their respective device. All 65 participants had used a desktop computer screen before and had considerable experience using it. On the contrary, only nineteen participants reported previous experiences of using a VR headset, and no one had used it more often than five times. Participants using the desktop computer screen did not need to adapt and learn how to use the device, while VR headset users did. Although we provided participants with clear explanations on how to use both devices and gave them 3 minutes to practice and familiarize with the virtual environment, it is possible that this was not enough time. VR headset users might have needed to spend cognitive effort on the handling of the VR device during the memorization task that the computer users did not need. The verbal comments of the test participants (cf. **Table 4**) indicate that the lack of experience with a VR headset device is one reason why the environment was considered less suitable for learning. This suggestion is in line with Makransky et al. (2019) and Polcar and Horejsi (2015) who also reported better learning performances for less immersive environments. Both suggested that lower learning might be linked to the lack of experience with the device. Makransky et al. (2019) explained that a lack of experience might lead to a higher cognitive load for participants using a VR environment.

A third explanation could be what we refer to as the "wow-effect." Using the VR headset was a new and impressive experience for most of the participants. For many of them, it was the first time they used such a high immersive device. Therefore, it can be assumed that they were distracted by the immersive experience and were, in consequence, less able to concentrate on the content provided in the virtual environment that they were supposed to memorize. For desktop users, however, this wow-effect was less pronounced as the device is less immersive and well known among university students. A similar argument was put forward by Makransky et al. (2019). They suggested that the strong hedonic component of the VR experience might lead participants to view the environment as playful and hence distract from the learning task. We were aware of the risk of a "wow-effect" and the potential impact it could have on learning. Therefore, participants in this study were asked to familiarize with the environment in a warm-up phase, what was expected to reduce initial excitement when experiencing for the first time such an immersive environment. It might be assumed however that the 3-min session was too short for the participants in our study. As VR becomes common and participants are expected to become more and more accustomed to its use, it could be speculated that the wow-effect as well the cognitive load due to lack of experience will be of lesser importance in the near future. Nonetheless, it might be interesting for this field of research to address questions regarding the wow-effect (e.g., does it wane over time, and how long does it last?) and the influence of extended usage experience (e.g., is cognitive load reduced in highly experienced VR users?), and their link to learning in future studies.

Finally, context-dependent memory might play a role in the unexpected effects reported in this piece of research. The essence of this well-established research topic (Godden and Baddeley, 1975; Herz, 1997; Smith and Vela, 2001) is that information is better recalled in the environment the information was learned in comparison to a new and unfamiliar context. Godden and Baddeley (1975) for example, found that divers learning underwater recalled the information better underwater than above. In the present study, both VR headset and desktop computer screen users answered the test (i.e., information recall) on the computer. As a consequence, computer users filled in the test on the same device they had been using while learning. In contrast, VR users answered on a different device, which might lead to reduced learning performances due to context-dependent memory. Although not explicitly mentioned in their article, Makransky et al. (2019) also asked their participants to answer the learning test on a desktop computer, regardless of which device they used for learning (personal communication). As they reported similar results to ours, context-dependent memory might have had an impact on the reduced learning performance of VR headset users in both studies. Future studies comparing learning with a VR headset with less immersive environments (e.g., desktop screen) should control for this potential error variable. This could be done either by presenting the assessment or test in the same technological environment as the learning took place (e.g., participants learning in VR take the test in VR as well) or by choosing an alternative medium (e.g., participants using high and low immersive devices complete the test with paper and pencil).

Overall, it is not possible to conclude with certainty that presence has no influence on learning based on the findings presented in this piece of research and the similar previous studies. It could be that that presence has a positive effect on learning, but this positive effect is outweighed by the negative consequences of using a highly immersive VR headset (i.e., cybersickness, increased cognitive workload, wow-effect or context dependent memory). In this regard, additional research is needed in which the confounding effect of these different influencing variables for learning can be controlled. This is important since the use of highly immersive VR headsets used to increase presence in this research context is always directly linked with an increase in the other influencing variables (e.g., increased cognitive workload). Possibilities to handle these issues have been discussed above: new VR systems need to be developed that reduce the prevalence of cybersickness, the recruitment of experienced participants might reduce the wow-effect, knowledge tests should be conducted in the same environment as the learning has taken place and cognitive workload needs to be held constant. One possibility to overcome these issues in experimental research might be to refrain from comparing VR systems with computer screen environments but to manipulate a VR environment which induces different levels of presence while keeping cognitive demand, cybersickness and other aspects constant.

With regard to learning, it was rather astonishing that the location did not show any influence. The manipulation check showed that the noise level in the cafeteria was considerably above the one in the lab (56 dB compared to 30 dB). Subjective ratings of participants' distraction and disturbance level also revealed significant differences between the two experimental conditions with the cafeteria being rated more disturbing and distracting than the lab. Since previous research has shown that noise (e.g., Cassidy and MacDonald, 2007) and interruptions (e.g., Altmann et al., 2014) have a negative effect on learning performance, it is intriguing that the location had no influence on learning in the present study. A possible explanation for this nil effect could be the sample of this study. Participants were students who are generally used to learn in noisy environments such as cafeterias and libraries. Accustomed to these places, their learning performance therefore might not have been influenced strongly enough. In addition, it could be speculated that the nil effects are a consequence of a participant behaviour particular in experimental studies. Although much effort has been put into making the experimental setup as natural as possible, participants were aware of the fact that they take part in a scientific study. This might have kept their concentration and focus on the presented content, while in a real learning situation, they might be more inclined to let themselves be distracted by the environment. This represents an extremely difficult challenge for future studies addressing the influence of the learning context experimentally. A possible approach could be a longitudinal field experiment, in which students' learning success with different systems is compared over a longer period of time–if possible, in a natural environment as implemented in this study.

Additional limitations relate to the generalisability of the results. It should be noted that only students were considered for this study. A generalisation across this population is therefore not appropriate. But as students are an important group of potential users of VR systems for learning, these results seem to be of practical relevance for an important part of the population. However, it would be of considerable interest for future studies to evaluate whether similar results (especially with regard to noisy environments) could be obtained with different user groups (e.g., pupils, adult trainees etc.). A second limitation concerns the controls used in the experiment. The controls in the VR headset were slightly easier to use compared to the ones for the desktop computer. VR headset users only had to use one controller (PowerMate) while desktop computer users had to deal with two (mouse and PowerMate). Four participants reported this in the comment section of the questionnaire as a factor that made them prefer the VR headset. This is particularly important since usability has been shown to be an important factor influencing the suitability of VR for learning (Hupont et al., 2015; Fernandes et al., 2016; Dolezal et al., 2017; Bryan et al., 2018; dela Cruz and Mendoza, 2018; Veronez et al., 2018). Although it would have been preferable to have used the same interaction mode for both experimental conditions in this study, this was technically not feasible since navigation *via* head movement was not applicable for the desktop computer condition. We would, however, expect the higher usability of the VR system to have a positive effect on learning performance and hence would expect an increase in the learning performance in the VR condition compared to the desktop computer condition—which was not the case. Another possible issue might have been the chosen time to adjust and practice to the novel device. Three minutes is possibly not enough time to familiarize participants with a novel device like a VR headset or at least not enough to eliminate a preference for the desktop computer due to experience. Extra time could have also reduced a possible wow-effect. Participants should have been given more time with the VR headset to overcome the wow and experience effect, although it remains unclear and subject for future research how much additional interaction time should be accorded to novice participants in order to overcome those two effects. Another way to control for the wow-effect would be running the experiment with experienced VR headset users. Lastly, while this study chose to conceptualize learning as memorization, other forms of learning (i.e., psychomotor or embodied learning) might be of interest for future research (Seo et al., 2018; Alvarez-Lopez et al., 2020). In addition, the VR environment in this study was simulating a place illusion in a pointandclick manner which is highly similar to a traditional screen usage, while plausibility illusion or agency in the VR world has not been assessed but might be of interest for future research (Slater, 2017; Gruber and Kaplan-Rakowski, 2020; Hurault et al., 2021).

5 CONCLUSION

Evidence of this study indicates that highly immersive VR may not always be beneficial for memorization, despite the higher level of perceived presence experienced by students using the VR headset and despite the fact that students are isolated from noisy and distracting environments. These results suggest that the link between presence and memorization with highly immersive devices is not as strong as often expected. Although performance was lower with the VR headset, users subjectively preferred the VR headset compared to a low-immersive desktop system. This might suggest that the VR headset may be a useful device for learning, nevertheless. This might especially be the case taking into consideration that feelings of presence increase students' motivation to learn (Psotka, 1995), suggesting that learning and presence might be positively correlated in the long run. Future studies should address the possible reasons for the reduced learning performance in VR (e.g., cybersickness, experience effect, wow effect, context-dependent memory) adopting a longitudinal research approach. This would provide in-depth knowledge of the VR learning processes and procedures that are necessary to create meaningful and useful learning environments for the future.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Internal Review Board of the Department of Psychology University of Fribourg. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

CO: Conceptualization, Methodology, Investigation, Formal analysis, Writing- Original draft preparation, Reviewing and Editing. AS: Conceptualization, Methodology, Writing-Reviewing and Editing.

ACKNOWLEDGMENTS

We want to thank Gabriela Lidia Hofmann, and Patrizia Baccarella, for their help with testing, as well as Huntley Thatcher for her many proofreads, and lastly, Kaspar Burger for his feedback on an earlier version of the manuscript.

REFERENCES

Abulrub, A. H. G., Attridge, A. N., and Williams, M. A. (2011). "Virtual Reality in Engineering Education: The Future of Creative Learning," in IEEE Global Engineeering Education Conference, Amman, Jordan, April 4–6, 2011 (EDUCON), 751–257. doi:10.1109/educon.2011.5773223

Ai-Lim Lee, E., Wong, K. W., and Fung, C. C. (2010). How Does Desktop Virtual Reality Enhance Learning Outcomes? A Structural Equation Modeling Approach. *Comput. Educ.* 55, 1424–1442. doi:10.1016/j.compedu.2010.06.006

Alhalabi, W. (2016). Virtual Reality Systems Enhance Students' Achievements in Engineering Education. *Behav. Inf. Techn.* 35, 919–925. doi:10.1080/0144929X.2016.1212931

Altmann, E. M., Trafton, J. G., and Hambrick, D. Z. (2014). Momentary Interruptions Can Derail the Train of Thought. *J. Exp. Psychol. Gen.* 143, 215–226. doi:10.1037/a0030986

Alvarez-Lopez, F., Maina, M. F., and Saigí-Rubió, F. (2020). Use of a Low-Cost Portable 3D Virtual Reality Gesture-Mediated Simulator for Training and Learning Basic Psychomotor Skills in Minimally Invasive Surgery: Development and Content Validity Study. *J. Med. Internet Res.* 22, e17491. doi:10.2196/17491

Alves Fernandes, L. M., Cruz Matos, G., Azevedo, D., Rodrigues Nunes, R., Paredes, H., Morgado, L., et al. (2016). Exploring Educational Immersive Videogames: an Empirical Study with a 3D Multimodal Interaction Prototype. *Behav. Inf. Techn.* 35, 907–918. doi:10.1080/0144929X.2016.1232754

Anderson, L. W., Krathwohl, D. R., and Bloom, B. S. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.

Baños, R. M., Guillen, V., Quero, S., García-Palacios, A., Alcaniz, M., and Botella, C. (2011). A Virtual Reality System for the Treatment of Stress-Related Disorders: A Preliminary Analysis of Efficacy Compared to a Standard Cognitive Behavioral Program. *Int. J. Human Comput. Stud.* 69, 602–613. doi:10.1016/j.ijhcs.2011.06.002

Beck, J. G., Palyo, S. A., Winer, E. H., Schwagler, B. E., and Ang, E. J. (2007). Virtual Reality Exposure Therapy for PTSD Symptoms after a Road Accident: An Uncontrolled Case Series. *Behav. Ther.* 38, 39–48. doi:10.1016/j.beth.2006.04.001

Biocca, F., and Delaney, B. (1995). "Communication in the Age of Virtual Reality," in *Virtual Reality as a Communication System* (Hillsdale, NJ, US: Lawrence Erlbaum Associates), 57–124.

Birrenbach, T., Zbinden, J., Papagiannakis, G., Exadaktylos, A. K., Müller, M., Hautz, W. E., et al. (2021). Effectiveness and Utility of Virtual Reality Simulation as an Educational Tool for Safe Performance of COVID-19 Diagnostics: Prospective, Randomized Pilot Trial. *JMIR Serious Games* 9, e29586. doi:10.2196/29586

Bles, W., Bos, J. E., de Graaf, B., Groen, E., and Wertheim, A. H. (1998). Motion Sickness: Only One Provocative Conflict? *Brain Res. Bull.* 47, 481–487. doi:10.1016/S0361-9230(98)00115-4

Bloom, B. S., Krathwohl, D. R., and Masia, B. B. (1984). *Bloom Taxonomy of Educational Objectives*. Copyright (c) by Pearson Education. Boston, MA: Allyn & Bacon. Available at: http://www.coun.uvic.ca/learn/program/hndouts/bloom.html.

Bowman, D. A., and McMahan, R. P. (2007). Virtual Reality: How Much Immersion Is Enough? *Computer* 40, 36–38. doi:10.1109/mc.2007.257

Bryan, S. J., Campbell, A., and Mangina, E. (2018). "Scenic Spheres - An AR/VR Educational Game," in 2018 IEEE Games, Entertainment, Media Conference (GEM), Galway, Ireland, August 15–17, 2018, 1–9. doi:10.1109/GEM.2018.8516456

Burdea, G., Richard, P., and Coiffet, P. (1996). Multimodal Virtual Reality : Input-Output Devices, System Integration, and Human Factors. *Int. J. Human Comput. Interaction,* 5–24.

Butler, J. (1997). *Excitable Speech: A Politics of the Performative*. New York: Psychology Press.

Buttussi, F., and Chittaro, L. (2018). Effects of Different Types of Virtual Reality Display on Presence and Learning in a Safety Training Scenario. *IEEE Trans. Vis. Comput. Graph.* 24, 1063–1076. doi:10.1109/TVCG.2017.2653117

Cadet, L. B., and Chainay, H. (2020). Memory of Virtual Experiences: Role of Immersion, Emotion and Sense of Presence. *Int. J. Human Comput. Stud.* 144, 102506. doi:10.1016/j.ijhcs.2020.102506

Carrozzino, M., and Bergamasco, M. (2010). Beyond Virtual Museums: Experiencing Immersive Virtual Reality in Real Museums. *J. Cult. Heritage* 11, 452–458. doi:10.1016/j.culher.2010.04.001

Cassidy, G., and MacDonald, R. A. R. (2007). The Effect of Background Music and Background Noise on the Task Performance of Introverts and Extraverts. *Psychol. Music* 35, 517–537. doi:10.1177/0305735607076444

Clarke, E. (2021). Virtual Reality Simulation-The Future of Orthopaedic Training? A Systematic Review and Narrative Analysis. *Adv. Simul* 6, 2. doi:10.1186/s41077-020-00153-x

Cryer, A., Kapellmann-Zafra, G., Abrego-Hernandez, S., Marin-Reyes, H., and French, R. (2019). "Advantages of Virtual Reality in the Teaching and Training of Radiation Protection during Interventions in Harsh Environments," in 2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Zaragoza, Spain, September 10-13, 2019 784–789. doi:10.1109/ETFA.2019.8869433

Davis, S., Nesbitt, K., and Nalivaiko, E. (2014). "A Systematic Review of Cybersickness," in Proceedings of the 2014 Conference on Interactive Entertainment, IE2014, Newcastle NSW, Australia, December 2-3, 2014 (New York, NY, USA: ACM), 8:1–8:9. doi:10.1145/2677758.2677780

Davis, T. R. V. (1984). The Influence of the Physical Environment in Offices. *AMR* 9, 271–283. doi:10.2307/25844010.5465/amr.1984.4277654

dela Cruz, D. R., and Mendoza, D. M. M. (2018). "Design and Development of Virtual Laboratory: A Solution to the Problem of Laboratory Setup and Management of Pneumatic Courses in Bulacan State University College of Engineering," in 2018 IEEE Games, Entertainment, Media Conference (GEM), Galway, Ireland, August 15-17, 2018 1–23. doi:10.1109/GEM.2018.8516467

Dolezal, M., Chmelik, J., and Liarokapis, F. (2017). "An Immersive Virtual Environment for Collaborative Geovisualization," in 2017 9th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games), Athens, Greece, October 05, 2017 (Athens, Greece: IEEE), 272–275. doi:10.1109/VS-GAMES.2017.8056613

Fisher, A. V., Godwin, K. E., and Seltman, H. (2014). Visual Environment, Attention Allocation, and Learning in Young Children. *Psychol. Sci.* 25, 1362–1370. doi:10.1177/0956797614533801

Freina, L., and Canessa, A. (2015). "Immersive vs Desktop Virtual Reality in Game Based Learning," in The European Conference on Games Based Learning (Academic Conference International Limited), 195.

Fried, C. B. (2008). In-class Laptop Use and its Effects on Student Learning. *Comput. Educ.* 50, 906–914. doi:10.1016/j.compedu.2006.09.006

Godden, D. R., and Baddeley, A. D. (1975). Context-dependent Memory in Two Natural Environments: on Land and Underwater. *Br. J. Psychol.* 66 (3), 325–331. doi:10.1111/j.2044-8295.1975.tb01468.x

Grassini, S., Laumann, K., and Rasmussen Skogstad, M. (2020). The Use of Virtual Reality Alone Does Not Promote Training Performance (But Sense of Presence Does). *Front. Psychol.* 11. doi:10.3389/fpsyg.2020.01743

Gruber, A., and Kaplan-Rakowski, R. (2020). "User Experience of Public Speaking Practice in Virtual Reality," in *Cognitive and Affective Perspectives on Immersive Technology*, 235–249. doi:10.4018/978-1-7998-3250-8.ch012

Haines, M. M., Stansfeld, S. A., Brentnall, S., Head, J., Berry, B., Jiggins, M., et al. (2001). The West London Schools Study: the Effects of Chronic Aircraft Noise Exposure on Child Health. *Psychol. Med.* 31, 1385–1396. doi:10.1017/S003329170100469X

Hayes, A. T., Straub, C. L., Dieker, L. A., Hughes, C. E., and Hynes, M. C. (2013). Ludic Learning. *IJGCMS* 5, 20–33. doi:10.4018/jgcms.2013040102

Heeter, C. (1992). Being There: The Subjective Experience of Presence. *Presence Teleop. Virtual Environ.* 1, 262–271. doi:10.1162/pres.1992.1.2.262

Herz, R. S. (1997). The Effects of Cue Distinctiveness on Odor-Based Context-dependent Memory. *Mem. Cogn.* 25, 375–380. doi:10.3758/bf03211293

Huang, Y. C., Backman, K. F., Backman, S. J., and Chang, L. L. (2016). Exploring the Implications of Virtual Reality Technology in Tourism Marketing: An Integrated Research Framework. *Int. J. Tourism Res.* 18, 116–128. doi:10.1002/jtr.2038

Hupont, I., Gracia, J., Sanagustin, L., and Gracia, M. A. (2015). "How Do New Visual Immersive Systems Influence Gaming QoE? A Use Case of Serious Gaming with Oculus Rift," in 2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX), Pilos, Greece, May 26-29, 2015 1–6. doi:10.1109/QoMEX.2015.7148110

Hurault, J.-C., Tedesco, A., and Brunel, L. (2021). " I Know what You Mean »: Investigating the Sense of agency in Learning an Abstract Mathematical Knowledge Using a Constructivist Method in Virtual Reality. *Lannee Psychol.* 121, 443–487. doi:10.3917/anpsy1.214.0443

Hutchinson, L. (2003). ABC of Learning and Teaching: Educational Environment. *BMJ: Br. Med. J.* 326, 810–812. doi:10.1136/bmj.326.7393.810

Jensen, L., and Konradsen, F. (2018). A Review of the Use of Virtual Reality Head-Mounted Displays in Education and Training. *Educ. Inf. Technol.* 23, 1515–1529. doi:10.1007/s10639-017-9676-0

Khan, A., and Rayner, G. D. (2003). Robustness to Non-normality of Common Tests for the many-sample Location Problem. *J. Appl. Maths. Decis. Sci.* 7, 187–206. doi:10.1155/s1173912603000178

Kozhevnikov, M., Gurlitt, J., and Kozhevnikov, M. (2013). Learning Relative Motion Concepts in Immersive and Non-immersive Virtual Environments. *J. Sci. Educ. Technol.* 22, 952–962. doi:10.1007/s10956-013-9441-0

LaViola, J. J. (2000). A Discussion of Cybersickness in Virtual Environments. *SIGCHI Bull.* 32, 47–56. doi:10.1145/333329.333344

Lee, E. A.-L., and Wong, K. W. (2014). Learning with Desktop Virtual Reality: Low Spatial Ability Learners Are More Positively Affected. *Comput. Educ.* 79, 49–58. doi:10.1016/j.compedu.2014.07.010

Lessiter, J., Freeman, J., Keogh, E., and Davidoff, J. (2001). A Cross-media Presence Questionnaire: The ITC-Sense of Presence Inventory. *Presence Teleop. Virtual Environ.* 10, 282–297. doi:10.1162/105474601300343612

Lohre, R., Bois, A. J., Athwal, G. S., and Goel, D. P. (2020). Improved Complex Skill Acquisition by Immersive Virtual Reality Training: A Randomized Controlled Trial. *JBJS* 102, e26. doi:10.2106/JBJS.19.00982

Lombard, M. (2016). A New Age of VR Involving All Five Senses [WWW Document]. International Society for Presence Research. Available at: https://ispr.info/2016/08/02/a-new-age-of-vr-involving-all-five-senses/ (Accessed 6 3, 2018).

Makransky, G., Terkildsen, T. S., and Mayer, R. E. (2019). Adding Immersive Virtual Reality to a Science Lab Simulation Causes More Presence but Less Learning. *Learn. Instruction* 60, 225–236. doi:10.1016/j.learninstruc.2017.12.007

Mayer, R. E. (1982). "Learning," in *Encyclopedia of Education Research*. Editor H. E. Mitzel (New York: Free Press), 1040–1058.

Mazloumi Gavgani, A., Walker, F. R., Hodgson, D. M., and Nalivaiko, E. (2018). A Comparative Study of Cybersickness during Exposure to Virtual Reality and "Classic" Motion Sickness: Are They Different? *J. Appl. Physiol.* 125, 1670–1680. doi:10.1152/japplphysiol.00338.2018

McCauley, M. E., and Sharkey, T. J. (1992). Cybersickness: Perception of Self-Motion in Virtual Environments. *Presence Teleop. Virtual Environ.* 1, 311–318. doi:10.1162/pres.1992.1.3.311

Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., and Davis, T. J. (2014). Effectiveness of Virtual Reality-Based Instruction on Students' Learning Outcomes in K-12 and Higher Education: A Meta-Analysis. *Comput. Educ.* 70, 29–40. doi:10.1016/j.compedu.2013.07.033

Mikropoulos, T. A., and Natsis, A. (2011). Educational Virtual Environments: A Ten-Year Review of Empirical Research (1999-2009). *Comput. Educ.* 56, 769–780. doi:10.1016/j.compedu.2010.10.020

Moreno, R., and Mayer, R. E. (2004). Personalized Messages that Promote Science Learning in Virtual Environments. *J. Educ. Psychol.* 96, 165–173. doi:10.1037/0022-0663.96.1.165

Moro, C., Štromberga, Z., and Stirling, A. (2017). Virtualisation Devices for Student Learning: Comparison between Desktop-Based (Oculus Rift) and mobile-based (Gear VR) Virtual Reality in Medical and Health Science Education. *Australas. J. Educ. Tec* 33. doi:10.14742/ajet.3840

Mousavia, M., Jenb, Y. H., Nurmaya, S., and Musac, B. (2013). A Review on Cybersickness and Usability in Virtual Environments. *Adv. Eng. Forum* 10, 34. doi:10.4028/www.scientific.net/AEF.10.34

Mütterlein, J. (2018). "The Three Pillars of Virtual Reality? Investigating the Roles of Immersion, Presence, and Interactivity," in 51st Hawaii International Conference on System Sciences (HICSS). doi:10.24251/HICSS.2018.174

Nichols, S., Cobb, S., and Wilson, J. R. (1997). Health and Safety Implications of Virtual Environments: Measurement Issues. *Presence Teleop. Virtual Environ.* 6, 667–675. doi:10.1162/pres.1997.6.6.667

Pan, Z., Cheok, A. D., Yang, H., Zhu, J., and Shi, J. (2006). Virtual Reality and Mixed Reality for Virtual Learning Environments. *Comput. Graph.* 30, 20–28. doi:10.1016/j.cag.2005.10.004

Parsons, T. D., and Rizzo, A. A. (2008). Affective Outcomes of Virtual Reality Exposure Therapy for Anxiety and Specific Phobias: A Meta-Analysis. *J. Behav. Ther. Exp. Psychiatry* 39, 250–261. doi:10.1016/j.jbtep.2007.07.007

Pears, M., and Konstantinidis, S. (2021). The Future of Immersive Technology in Global Surgery Education. *Indian J. Surg.*, 1–5. doi:10.1007/s12262-021-02998-6

Piedra, J. A., Ojeda-Castelo, J. J., Quero-Valenzuela, F., and Piedra-Fdez, I. (2016). "Virtual Environment for the Training of the Hands in Minimally Invasive Thoracic Surgery," in 2016 8th International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES), 1–4. doi:10.1109/VS-GAMES.2016.7590375

Polcar, J., and Horejsi, P. (2015). Knowledge Acquisition and Cyber Sickness: a Comparison of Vr Devices in Virtual Tours. *MM Sci. J.* 2015, 613–616. doi:10.17973/MMSJ.2015_06_201516

Potdevin, D., Clavel, C., and Sabouret, N. (2021). A Virtual Tourist Counselor Expressing Intimacy Behaviors: A New Perspective to Create Emotion in Visitors and Offer Them a Better User Experience? *Int. J. Human Comput. Stud.* 150, 102612. doi:10.1016/j.ijhcs.2021.102612

Price, M., and Anderson, P. (2007). The Role of Presence in Virtual Reality Exposure Therapy. *J. Anxiety Disord.* 21, 742–751. doi:10.1016/j.janxdis.2006.11.002

Psotka, J. (1995). Immersive Training Systems: Virtual Reality and Education and Training. *Instr. Sci.* 23, 405–431. doi:10.1007/bf00896880

Radianti, J., Majchrzak, T. A., Fromm, J., and Wohlgenannt, I. (2020). A Systematic Review of Immersive Virtual Reality Applications for Higher Education: Design Elements, Lessons Learned, and Research Agenda. *Comput. Educ.* 147, 103778. doi:10.1016/j.compedu.2019.103778

Rahm, S., Wieser, K., Wicki, I., Holenstein, L., Fucentese, S. F., and Gerber, C. (2016). Performance of Medical Students on a Virtual Reality Simulator for Knee Arthroscopy: an Analysis of Learning Curves and Predictors of Performance. *BMC Surg.* 16, 14. doi:10.1186/s12893-016-0129-2

Rasheed, F., Onkar, P., and Narula, M. (2015). "Immersive Virtual Reality to Enhance the Spatial Awareness of Students," in Proceedings of the 7th International Conference on HCI, IndiaHCI 2015, Guiwahati, India, December 17-19, 2015 (New York, NY, USA: Association for Computing Machinery), 154–160. doi:10.1145/2835966.2836288

Ray, A. B., and Deb, S. (2016). "Smartphone Based Virtual Reality Systems in Classroom Teaching - A Study on the Effects of Learning Outcome," in 2016 IEEE Eighth International Conference on Technology for Education (T4E), Mumbai, India, January 16, 2017 68–71. doi:10.1109/T4E.2016.022

Riva, G., Mantovani, F., Capideville, C. S., Preziosa, A., Morganti, F., Villani, D., et al. (2007). Affective Interactions Using Virtual Reality: The Link between Presence and Emotions. *Cyberpsychol. Behav.* 10, 45–56. doi:10.1089/cpb.2006.9993

Rizzo, A., Parsons, T. D., Lange, B., Kenny, P., Buckwalter, J. G., Rothbaum, B., et al. (2011). Virtual Reality Goes to War: A Brief Review of the Future of Military Behavioral Healthcare. *J. Clin. Psychol. Med. Settings* 18, 176–187. doi:10.1007/s10880-011-9247-2

Salzman, M. C., Dede, C., Loftin, R. B., and Chen, J. (1995). The Design and Evaluation of Virtual Reality-Based Learning Environments. *Presence Teleop. Virtual Environ.*, 293–316. (special issue on education).

Salzman, M. C., Dede, C., Loftin, R. B., and Chen, J. (1999). A Model for Understanding How Virtual Reality Aids Complex Conceptual Learning. *Presence Teleop. Virtual Environ.* 8, 293–316. doi:10.1162/105474699566242

Schofield, D. (2012). "Mass Effect: A Chemical Engineering Education Application of Virtual Reality Simulator Technology. 8 (1), 17

Seo, J. H., Smith, B. M., Cook, M., Malone, E., Pine, M., Leal, S., et al. (2018). "Anatomy Builder VR: Applying a Constructive Learning Method in the Virtual Reality Canine Skeletal System," in *Advances in Human Factors in Training, Education, and Learning Sciences, Advances in Intelligent Systems and Computing*. Editor T. Andre (Cham: Springer International Publishing), 245–252. doi:10.1007/978-3-319-60018-5_24

Sharar, S. R., Miller, W., Teeley, A., Soltani, M., Hoffman, H. G., Jensen, M. P., et al. (2008). Applications of Virtual Reality for Pain Management in Burn-Injured Patients. *Expert Rev. Neurotherapeutics* 8, 1667–1674. doi:10.1586/14737175.8.11.1667

Sharples, S., Cobb, S., Moody, A., and Wilson, J. R. (2008). Virtual Reality Induced Symptoms and Effects (VRISE): Comparison of Head Mounted Display (HMD), Desktop and Projection Display Systems. *Displays* 29, 58–69. doi:10.1016/j.displa.2007.09.005

Shu, Y., Huang, Y.-Z., Chang, S.-H., and Chen, M.-Y. (2019). Do virtual Reality Head-Mounted Displays Make a Difference? A Comparison of Presence and Self-Efficacy between Head-Mounted Displays and Desktop Computer-Facilitated Virtual Environments. *Virtual Reality* 23, 437–446. doi:10.1007/s10055-018-0376-x

Skarredghost (2017). *The Sense of Taste in Virtual Reality*. The Ghost Howls. Available at: https://skarredghost.com/2017/07/07/sense-taste-virtual-reality/ (Accessed 6 3, 2018).

Slater, M., and Wilbur, S. (1997). A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. *Presence Teleop. Virtual Environ.* 6, 603–616. doi:10.1162/pres.1997.6.6.603

Slater, M. (2003). A Note on Presence Terminology. *Presence Connect.* 3, 1–5.

Slater, M. (2017). "Implicit Learning through Embodiment in Immersive Virtual Reality," in *Virtual, Augmented, and Mixed Realities in Education, Smart Computing and Intelligence*. Editors D. Liu, C. Dede, R. Huang, and J. Richards (Singapore: Springer), 19–33. doi:10.1007/978-981-10-5490-7_2

Smith, S. M., and Vela, E. (2001). Environmental Context-dependent Memory: A Review and Meta-Analysis. *Psychon. Bull. Rev.* 8, 203–220. doi:10.3758/bf03196157

Spira, J. B., and Feintuch, J. B. (2005). *The Cost of Not Paying Attention: How Interruptions Impact Knowledge Worker Productivity*. New York, NY: Report from Basex.

Stevens, J. A., and Kincaid, J. P. (2015). The Relationship between Presence and Performance in Virtual Simulation Training. *OJMSi* 03, 41–48. doi:10.4236/ojmsi.2015.32005

Sutcliffe, A., Gault, B., and Shin, J.-E. (2005). Presence, Memory and Interaction in Virtual Environments. *Int. J. Human Comput. Stud.* 62, 307–327. doi:10.1016/j.ijhcs.2004.11.010

Sylaiou, S., Mania, K., Karoulis, A., and White, M. (2010). Exploring the Relationship between Presence and Enjoyment in a Virtual Museum. *Int. J. Human Comput. Stud.* 68, 243–253. doi:10.1016/j.ijhcs.2009.11.002

Veronez, M. R., Gonzaga, L., Bordin, F., Kupssinsku, L., Kannenberg, G. L., Duarte, T., et al. (2018). "RIDERS: Road Inspection & Driver Simulation," in 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Tuebingen/Reutlingen, Germany, August 18-22, 2018 715–716. doi:10.1109/VR.2018.8446207

Vogel, J. J., Greenwood-Ericksen, A., Cannon-Bowers, J., and Bowers, C. A. (2006). Using Virtual Reality with and without Gaming Attributes for Academic Achievement. *J. Res. Techn. Educ.* 39, 105–118. doi:10.1080/15391523.2006.10782475

Weech, S., Kenny, S., and Barnett-Cowan, M. (2019). Presence and Cybersickness in Virtual Reality Are Negatively Related: A Review. *Front. Psychol.* 10, 158. doi:10.3389/fpsyg.2019.00158

Winter, J., Cotton, D., Gavin, J., and Yorke, J. D. (2010). Effective E-Learning? Multi-Tasking, Distractions and Boundary Management by Graduate Students in an Online Environment. *ALT-J* 18, 71–83. doi:10.1080/09687761003657598

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

Wyon, D. P. (2004). The Effects of Indoor Air Quality on Performance and Productivity. *Indoor Air* 14, 92–101. doi:10.1111/j.1600-0668.2004.00278.x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

ORIGINAL RESEARCH
published: 14 April 2022
doi: 10.3389/frvir.2022.810797

Check for updates

Informing and Evaluating Educational Applications With the Kirkpatrick Model in Virtual Environments: Using a Virtual Human Scenario to Measure Communication Skills Behavior Change

Stephanie Carnell [1], Alexandre Gomes De Siqueira [2], Anna Miles [3] and Benjamin Lok [2]*

[1]Department of Computer Science, University of Central Florida, Orlando, FL, United States, [2]Department of Computer and Information Sciences and Engineering, University of Florida, Orlando, FL, United States, [3]Speech Science, The University of Auckland, Auckland, New Zealand

OPEN ACCESS

Edited by:
Missie Smith,
Independent researcher, Detroit, Michigan, United States

Reviewed by:
Simon Hoermann,
University of Canterbury, New Zealand
Meredith Carroll,
Florida Institute of Technology, United States

Correspondence:
Stephanie Carnell
stephaniecarnell@gmail.com

Specialty section:
This article was submitted to Virtual Reality and Human Behaviour, a section of the journal Frontiers in Virtual Reality

Received: *07 November 2021*
Accepted: *28 March 2022*
Published: *14 April 2022*

Citation:
Carnell S, Gomes De Siqueira A, Miles A and Lok B (2022) Informing and Evaluating Educational Applications With the Kirkpatrick Model in Virtual Environments: Using a Virtual Human Scenario to Measure Communication Skills Behavior Change. Front. Virtual Real. 3:810797. doi: 10.3389/frvir.2022.810797

Increasingly, virtual environments are being used in educational and training applications. As with other types of applications that use virtual environments, these scenarios must be evaluated in terms of user experience. However, they also should be evaluated on the efficacy of the training or learning provided, so as to ensure learning transfer. Frameworks, such as the Kirkpatrick Model, exist to evaluate training scenarios, but application of these frameworks has not been fully utilized in development of virtual environment-based education and training. To address this gap and to also share our process with other virtual environment developers, we discuss our experience applying the Kirkpatrick Model to an existing virtual human (VH) application for medical communication skills training. The Kirkpatrick Model provides different levels of evaluation for training programs that include learners' reactions to the training, the knowledge acquired from the training, behaviors indicating the training was applied, and the degree high-level results were impacted as a result of the training. While we discuss all of the Model's levels, our focus for this work is Level 3 Behavior. The Kirkpatrick Model currently recommends that behavioral change may only be measured while a trainee is working in a real-world context. However, given existing evidence that VH applications have been shown to elicit real-world behaviors from participants, we suggest that VH training scenarios may be a method of measuring Behavior level metrics before trainees are evaluated *in situ*. Initial support for this suggestion is provided by our study examining whether VHs can elicit changes in communication skills learners' message production behavior over time. This study indicates that learners displayed changes in several metrics over the course of the semester. Based on this finding, we suggest a direction for future research: observing learner behavior in a virtual environment as a pre-cursor to behavioral measures while in a real-world scenario.

Keywords: virtual reality, educational technology, communication skills, kirkpatrick model, virtual humans, virtual patients

1 INTRODUCTION

Increasingly, virtual environments are being used in educational and training applications in a variety of scenarios, such as medical training and interpersonal communication skill building (Xie et al., 2021). Thus, in addition to general concerns regarding the user experience of these applications, virtual environment developers must often also evaluate the training provided by these scenarios. However, evaluating this training aspect of virtual environments can be challenging: in a recent review of virtual reality (VR) applications for skills training, Xie et al. note that identifying the particular factors of training that should be targeted to ensure learning transfer is difficult (Xie et al., 2021). The authors also note the existence of specific concerns regarding learning transfer between virtual training and the real world (Xie et al., 2021).

In the medical domain, researchers have applied the Kirkpatrick Model to address this question of how to evaluate educational virtual environments to target learning transfer (see Zaveri et al. (2016); Kundhal and Grantcharov (2009); Beal et al. (2017); Delisle et al. (2019) for examples). The Kirkpatrick Model is often considered the gold standard for evaluation of training but has not yet been widely applied in virtual environment-based training. The Kirkpatrick Model has four different "levels" by which an training program may be evaluated: 1) Reaction, 2) Learning, 3) Behavior, and 4) Results. The Kirkpatrick Model is suitable for evaluating virtual environment-based education and training across a range of criteria: the Model includes typical user experience measures at the first (Reaction) level, such as engagement or satisfaction, while also providing guidance to identify measures focused on the learning outcomes of the training program. To identify these outcome-focused measures, the Kirkpatrick Model recommends beginning with the final level, the Results level, to ensure a training program meets an organization's larger mission or purpose. Once potential measures for the results level are identified, training program developers can work backward from this larger vision to identify behaviors and skills that should be targeted in the training program itself.

In this paper, we recommend the application of the Kirkpatrick Model to educational and training applications using virtual environments and explain our process of applying the Kirkpatrick Model to virtual human (VH) healthcare communication skills training. While other medical education research has applied the Kirkpatrick Model for training program evaluation, we detail our process here so that developers of non-medical virtual environments or simulations may benefit from the best practices of the medical education community. Our work here describes the process of applying the Kirkpatrick Model to a desktop virtual human (VH) application for healthcare students' communication skills. By using this process, we were able to identify relevant metrics that allowed us to evaluate whether VHs can elicit changes in healthcare students' communication skills over time.

A brief overview of our process is as follows: we began by identifying an existing problem in healthcare communication, patient adherence. Patient adherence refers to the level patients follow the medical instructions given to them by their healthcare providers. While we did not measure patient adherence directly, identifying patient adherence as a Results level measure directly informed our lower-level measures, as suggested by the Kirkpatrick Model. By aiming to improve patient adherence, we were then able to identify healthcare provider behaviors to target in our application–those that promote higher patient adherence. We then identified six metrics related to a healthcare provider's message production behavior, or how one transforms one's thoughts into messages to communicate with others. While our focus for this work is primarily discussing the Behavior level of the Kirkpatrick Model, we also detail potential level two and level one measures for our communication skills training application.

In addition to discussing our application of the Kirkpatrick Model, we also suggest that educational VH scenarios may be applicable to several levels of the Kirkpatrick Model. The latest version of the Model states that Behavior measures may only be evaluated when learners apply training in real-world settings. However, as VHs can elicit real-world behaviors from participants (Cassell et al., 2009; Kleinsmith et al., 2015), we suggest that evaluations of behaviors may begin to be examined with VHs by using behavioral measures that can be used in both the virtual and real worlds. In other words, developers may be able to gain insight regarding the efficacy of the virtual environment training by incorporating behavioral measures in the VH training itself, potentially lessening the gap between the Learning and Behavior levels of the Model.

To add to the existing literature that suggests that VHs can elicit real-world behaviors, we present our study examining whether VHs can elicit changes in communication skills learners' message production over time. For this study, we invited speech-language pathology students to interview two virtual patients (VPs) over the course of their academic semester. Using the Kirkpatrick Model, we identified six message production metrics that to target patient adherence, or the degree to which a patient follows their providers' healthcare instructions. Using the VP interview data, we compared students' message production at different points in their academic semester using these message production metrics. This study indicates that learners displayed changes in several metrics over the course of the semester, thus suggesting the potential for VHs to capture trainee behavioral data.

2 RELATED WORK

In this section, we discuss previous applications of the Kirkpatrick Model to virtual environments for education and/or training. We briefly introduce the model in **Section 2.1** for discussion purposes, but a fuller description of the Model and each level is provided in **Section 3.2**. We note that for brevity, when we discuss the Model, we refer to the New World Kirkpatrick Model, as presented in Kirkpatrick and Kirkpatrick's 2016 book (Kirkpatrick and Kirkpatrick, 2016). This New World Model is the latest iteration of the Kirkpatrick Model first presented in Dr. Kirkpatrick's dissertation in 1954 (Kirkpatrick, 1954). In **Section**

2.2, we also discuss existing methods for evaluating healthcare communication skills training while in virtual environments, as medical communication skills training is the educational domain of interest for our VH scenario.

It is important to note that systems deploying virtual environments may cover systems with a variety of characteristics, ranging from low-to high-tech and from fully immersive environments that require the use of head-mounted displays (HMDs) to non-fully immersive 2D VR systems administered without HMDs (Li et al., 2011). While our work with VHs is focused on non-fully immersive systems, we believe that the Kirkpatrick concepts discussed are applicable to both fully- and non-fully immersive systems, as the Kirkpatrick Model is not reliant on any particular type of training in order to be used. Consequently, in this section, we discuss a number of systems that range in terms of immersiveness.

2.1 Existing Applications of the Kirkpatrick Model to Training in Virtual Environments

The Kirkpatrick Model has four levels for evaluating training (Kirkpatrick and Craig, 1970):

- Level 4 Results—the degree targeted outcomes occur
- Level 3 Behavior—the degree participants apply concepts learned in training
- Level 2 Learning—the degree participants acquire intended knowledge in training
- Level 1 Reaction—the degree participants find the training favorable

Results–the highest level and the "ultimate" outcome of the training scenario–measure the impact of the training on the organization level (productivity gains, cost savings, employee attitude/morale) (Brogden and Taylor, 1950). The Kirkpatrick Model advocates evaluating training scenarios with the results level in mind first, so that the impact of these results may inform the lower levels. Behavior, Level 3, measures the degree individuals actually use what they learned in training when they are on the job (Alliger et al., 1997). The next level, Learning (Level 2), is defined in this context as knowledge, skills and feelings acquired in the short term at the end of training (the simplest and most commonly used measurement) and in the long term to assess retention of what was learned. Reaction (Level 1) is a measurement of trainees' feelings toward the training program in terms of utility and enjoyment, and it is the most commonly collected type of evaluation data (Bassi et al., 1996).

Given the popularity of the Kirkpatrick Model, several applications using virtual environments have applied the Model, but the use of the full-breadth of evaluation levels appears to be rare. The majority of studies on virtual environment-based training have reported positive results regarding users' reactions (Level 1) (Schmidt and Stewart, 2009; Alaraj et al., 2011; Loukas et al., 2011; Kidd et al., 2012; Cohen et al., 2013). However, fewer studies have attempted to reach Levels 3 and 4. For example, Suàrez et al. applied the Kirkpatrick Model as a framework to compare learning with virtual human role-players and a variety of other training methods, including real human role-players (Suárez et al., 2021). The authors note that they only focused on Levels 1 and 2 of the Model explicitly because the higher levels can only be evaluated "once a long period of time has elapsed after training" (Suárez et al., 2021). While certain aspects of Levels 3 and 4, such as monitoring learners' on-the-job behavior, do require some time to pass, the important behaviors to target in Level 3 can potentially be incorporated into educational virtual environments to begin understanding the impact of the training, as we will discuss in **Section 3.2**.

Similarly, Grabowski et al. developed a virtual reality-based pilot training simulation for underground coal miners (Grabowski and Jankowski, 2015). Work in the mining industry has been described as dirty, dark, wet, noisy, hot, uncomfortable and as being one of the most dangerous industries (Van Wyk and De Villiers, 2009), supporting the idea that the Kirkpatrick Model is a suitable evaluation method toward reducing the gap between theoretical training and practice. In this context, Grabowski et al. applied a training questionnaire based on the Kirkpatrick Model. Similar to many other educational applications using virtual environments, they focused on evaluating lower levels—in this case Level 1 (Reactions)—with less emphasis on Levels 2, 3, and 4.

As another example from the healthcare context, Zaveri et al. compared an online learning platform and a virtual human-based module (on Second Life) simulating pediatric sedation procedures (Zaveri et al., 2016). In contrast to the previously described research applying the Kirkpatrick Model, the authors attempted to evaluate their work regarding the first three of Kirkpatrick's levels. The results showed positive findings for Kirkpatrick's Level 1 (participants had a positive reaction to the experience). However, no statistically significant differences were found regarding Levels 2 and 3 when comparing the virtual-human module and the baseline web-based module. In another noteworthy example, Kundhal, et al. compared performance in a virtual environment to actual operating room performance by applying a checklist, effectively evaluating Levels 3 and 4 of the Model (Kundhal and Grantcharov, 2009). This work demonstrated that training in virtual environments can impact those two levels when simulating real environments.

Taken together, these efforts suggest a trend toward applying the Kirkpatrick Model to educational and training virtual environments, with more frequent application of the full Model in the healthcare education context. However, little is mentioned for these applications about how the Model was applied and to what extent it was used beyond questionnaires for the evaluation phase of those studies. Our work contributes to the field of virtual environments for education training by describing how we adapted the Model to virtual environment training, specifically in the context of a healthcare scenario, and how the Kirkpatrick Model can help virtual environment developers and researchers plan the overall goals and metrics for their proposed systems.

2.2 Educational Measures in Applications for Healthcare Communication Skills Training in Virtual Environments

Given the importance of doctor-patient communication, researchers in healthcare education have developed approaches to measure students' communication competency in real environments. One such approach is the Control, Explaining, Listening and Influencing (CELI) model (Wouda et al., 2011), which aims to promote patient-centered communication. Given the complex nature of patient-centered communication, the developers of this model note that many healthcare communications skill training scenarios suffer from a mismatch between the learning objectives and skills taught in the training. To better address this mismatch, the CELI model was developed.

The existence of the CELI model for real world competency measurement would suggest that a simple method to address communication skills measurement in virtual environments is to use the CELI model in a virtual environment itself. However, follow-up research using the CELI model in the real world revealed that healthcare students require deliberate practice in order to improve communication skills past a "satisfactory" level (Wouda and van de Wiel, 2012). Deliberate practice involves a learner engaging in activities with explicit learning goals that can allow the learner to challenge any behaviors that are unconscious and sub-optimal. According to Wouda and van de Wiel, the key components of deliberate practice for healthcare communication skills training are as follows (Wouda and van de Wiel, 2013):

- "Learning tasks with well-defined goals"
- "Stimulating learning tasks of short duration with opportunities for immediate feedback, reflection, and corrections"
- "Having ample opportunities for repetition, gradual refinements, and practice in challenging situations"
- "Being motivated to improve"

Characteristics such as the "well-defined goals" and the need for tasks with "short duration" and "immediate feedback" suggest a narrower scope than a holistic view of patient-centered communication, which is the aim of the CELI model. Thus, from the existing literature on measuring healthcare communication competency, we see two important goals that should be addressed by healthcare communication skills scenarios: 1) alignment between a scenario's stated learning objectives and the skills being taught 2) a narrower scope than broadly improving patient-centered communication.

Evidence of the latter goal is present in many healthcare communication skills training scenarios, as many of these scenarios focus on specific skills or types of communication. For example, several virtual patient scenarios focus on developing student empathy (Halan et al., 2015; Foster et al., 2016). Other applications have focused on information discovery, notably the Virtual People Factory system (discussed further in **Section 3.1**), the existing system to which we applied the Kirkpatrick Model in this work. Still other non-fully immersive systems, such as SIDNIE, targeted specific communication skills needed for working with a particular group of patients. In the case of SIDNIE, the system targeted learners' unbiased and age-appropriate language when interacting with pediatric patients (Dukes et al., 2013).

These systems use a variety of methods specific to the communication skill of interest to measure learners' performance. For example, information discovery in VPF2 is measured by students' discovery of pre-defined pieces of important diagnostic information. Similarly, choosing the more unbiased and age-appropriate questions built into SIDNIE yields better performance. On the other hand, the empathy systems have used simulations to collect learner communication skills behavior that is then later evaluated by an expert grader using an existing framework, such as the Empathic Communication Coding System for empathy. While there are clearly a variety of methods to measure communications skills in virtual environments, common to many of these methods is the incorporation of an expert in healthcare communication to provide guidance on metric development. However, the process for working with these experts is often not explicitly discussed, especially in terms of ensuring alignment between a scenario's learning objectives and the skills being taught. The Kirkpatrick Model is a good candidate for a framework to address these concerns and may provide a common perspective by which to discuss and compare these different metrics for healthcare communication skills training, despite originating from different skills and being applied to different virtual environments.

3 THE KIRKPATRICK MODEL AND ITS APPLICATION TO AN EDUCATIONAL VIRTUAL HUMAN HEALTHCARE SCENARIO

We now discuss our application of the Kirkpatrick Model to a specific educational context: medical communication skills training. Since we were applying the Model to an existing training scenario, we begin this section with details of Virtual People Factory, a desktop-based system that features conversational VHs (see **Section 3.1**). Then, as the Model proposes addressing the highest level (Level 4 Results) first in order to address the learning-practice gap, we describe our process in **Section 3.2** with details of Level 4 and work downward.

3.1 Virtual People Factory 2.0

VPF2 is a non-fully immersive application accessible online that enables creation of and interaction with VHs. It is an iteration of the conversational modeling system, Virtual People Factory, developed by Brent Rossen in (Rossen, 2011). VPF2 was designed to allow individuals without technical expertise but with a particular domain expertise, such as a healthcare instructor, to author a VH that can then be interviewed in the same application. The VPF2 authoring process mostly focuses on the creation of the VH script, which contains the dialogue

FIGURE 1 | A screenshot of VPF2's chat interface with Lilly Smith.

responses a VH can provide, as well as the corresponding questions that can elicit those dialogue responses.

Script authors can also define various meta-data, such as discoveries and topics, for a VH script. A discovery is an important piece of information that should be uncovered by the learner over the course of a VH interview. Example discoveries from a medical VH include "Difficulty with tough foods" and "Coughs while eating." In the context of a medical interview, these discoveries may be important for making a diagnosis. Another meta-data option provided by VPF2 is topics. Topics can be used to group question and response pairs. For instance, a virtual patient's script could contain the question "How does your swallowing problem affect your social life?" under the topic of "Chief Complaint."

In addition to the virtual human authoring capabilities provided by VPF2, the application also enables the interviewing of VHs in an online interface. This interface allows remote VH interviewers to ask a VH questions while using a personal desktop or laptop device. Typically, the interview is conducted in a chat interaction style, as shown in **Figure 1**: interviewers can type questions into an input box, and VPF2 will match the typed question to the available phrasings in the virtual human script. If a matching phrasing is found, the VH responds with the corresponding script response. If no matching phrasing is found, a standard exception response ("Sorry, I don't understand what you just said. Can you say it another way?") is returned instead. If an interviewer asks a question that is mapped to a discovery, that discovery is considered "uncovered" and is counted toward an interviewer's discovery score. A discovery score is calculated by dividing the total number of uncovered discoveries by the total discoveries in a VH scenario.

We place our work on the reality-virtuality continuum (?) by describing VPF2 as a system to create and interact with conversational VHs. Conversational VHs combine the virtual components of a conversation partner, such as speech, gestures, animations, virtual characters, and varying capabilities to understand the user's verbal and nonverbal inputs. The conversational virtual humans exist on a continuum of levels of immersion from displays such as on mobile phones and laptops to immersive displays such as head-mounted displays and CAVE-like systems. The conversational VHs discussed in this paper included VHs capable of conversational dialogue (either typed or spoken) restricted to the topic domain and deployed on lower-immersion laptop and desktop displays. This form factor was chosen to enable an educational experience that could be integrated into an existing curriculum and accessible via the resources available to the enrolled students.

3.2 The Kirkpatrick Model Applied

Our stakeholder for this work (listed as the third author, AM) was interested in integrating existing virtual patients (VPs) into a clinical practicum course for speech-language therapy (SLT) students. The clinical practicum course is part of the students' clinical training, and the VP interviews were integrated into the course to provide support for the students' final clinical exam. An overview of our process for improving the impact of VP training and the takeaways we identified from each step is as follows:

- Work with educators to determine and prioritize the most important results to target → patient adherence to healthcare recommendations and patient-centered communication
- Work with educators to determine and prioritize the most important behaviors learners need to exhibit to impact the above outcomes → SLTs should use language that promotes patient adherence and patient-centered communication
- Work with educators to clearly map VP learning objectives to the expected behaviors → VPs should recognize and reward learners' language that promotes patient adherence and patient-centered communication
- Work with learners to understand their reaction to the training → learners should find the VP scenario useful in practicing using language that promotes patient adherence and patient-centered communication

3.2.1 Results

First, we began by discussing with our stakeholder the planned high-level results we wished to target using the existing VPs. Our discussion centered on which problems in healthcare might be impacted by a healthcare providers' communication skills. (In our case, as we had an existing VPs focused on communication skills, this topic framed our initial results discussion, but if one is creating entirely new educational virtual environment application, this discussion will likely be more open-ended.) One pressing issue that arose in these initial discussions was that of patient adherence. Patient adherence, or the ability to follow a provider's instructions for care, has been linked to successful patient outcomes. For example, for patients at risk of heart disease, patient non-adherence can greatly influence survival rates (Martin et al., 2005). In addition to the health risks associate with non-adherence, there is also a great economic cost. A 2004 survey estimated the "monetary waste" in the United States associated with non-adherence could be as great as $300 billion per year (DiMatteo, 2004). In 2005, the cost of medical non-adherence alone was calculated to approximately $100 billion annually.

Also of interest to our stakeholder was the applications's cultivation of students' holistic interviewing skills, an important aspect of patient-centered communication. Patient-centered communication is a method of communicating with patients that promotes a holistic understanding of patients rather than a sole focus on the patient's medical problem, so an important skill healthcare students should cultivate is asking questions on biomedical topics and social topics. While holistic interviewing is often targeted as a result on its own, a lack of patient-centered communication may contribute to a lack of patient adherence. Take as an example the medical case of focus for our VH simulation, dysphagia. Dysphagia is characterized by difficulty swallowing, so an important factor to discuss is the patient's diet: what types of food they eat, the hardness/softness of these foods, and so on. The importance of food in dysphagia management makes gathering a holistic perspective of the patient especially critical, as dysphagia patients' cultures and food can have a large impact on their medical condition (Dikeman and Riquelme, 2002). However, such details about patients' culture and food practices may not arise without the provider attempting to uncover a holistic view of the patient.

3.2.2 Behaviors

After identifying improved patient adherence and holistic interviewing as important outcomes to target, our next step was to identify healthcare providers' communication behaviors that could affect these outcomes. The Kirkpatrick Model states that Level 3 Behaviors can only truly be evaluated when learners apply their training in the corresponding real-world scenario (Kirkpatrick and Kirkpatrick, 2016). Additionally, these behaviors should also be evaluated over the course of weeks or months after the training to ensure that the training is effective.

However, VHs have been to shown to elicit real-world behaviors from humans in a variety of situations, such as the presence of public speaking anxiety (Slater et al., 2006), context-switching in child peer-to-peer communication (Cassell et al., 2009), and display of empathy with virtual patients (Kleinsmith et al., 2015). These behaviors are often demonstrated despite "low representational and behavioral fidelity" (Slater et al., 2006) or even acknowledgement from participants that the VH was "less authentic" than an interaction with a real patient (Raij et al., 2006; Kleinsmith et al., 2015). Thus, we see that even less immersive systems can elicit real world behaviors with VHs. Similarly, VR has also been used to study psychological phenomenon previously only studied in physical settings (Fox et al., 2009).

Based on VHs' abilities to elicit realistic behaviors from users, we suggest that the Kirkpatrick Model's Level 3 Behaviors may also be observable in our application as well. In other words, we may begin to observe learners' behavior during training (the VH simulation) itself and may use the training as a method to measure learner behavior over time. Being able to observe such measures early while learners are still interacting with VHs may give developers insight as to whether the proper behaviors are being learned from the training.

For our particular application, we should therefore identify behaviors related to patient adherence that are meaningful and measurable in both the real world and the virtual world that are observable over time. Research in doctor-patient communication indicates that cognitive factors, mostly the patient's ability to understand medical information, are central to issues of patient adherence (Martin et al., 2005). Two popular recommendations for communication behaviors that promote patient understanding are reducing the use of medical jargon (Martin et al., 2005; Graham and Brookey, 2008; Oates and Paasche-Orlow, 2009; Green et al., 2014; Speer, 2015) and using simple language (Oates and Paasche-Orlow, 2009; Green et al., 2014; Speer, 2015). While these recommendations may seem simple, failure to follow them can have severe consequences. Patients have expressed concerns about providers who fail to use these recommended behaviors (Waisman et al., 2003; Shaw et al., 2009). In some cases, failure to practice these strategies has also led to malpractice lawsuits (Gordon, 1996).

A VH training scenario for medical communication skills can therefore use measures related to reduced medical jargon and simple language to promote patient adherence, but exactly how these how these measures should be calculated is not necessarily obvious. For example, easily calculable measures exist to calculate language complexity, such as the Flesch Reading Ease (Flesch, 1948), but how should these measures be applied to a learner's communication behavior in a VH scenario? Should the goal be simply to promote that language complexity should be as low as possible? Additionally, we should also consider how these measures might relate to measures for patient-centered communication, such as asking the patient questions about relevant social or cultural topics.

At this stage, we recommend working with stakeholders to identify a framework to unify the behaviors of interest for the virtual environment scenario, as the framework can assist in further refining how the measures ought to be defined. In our work, our focus on specific communication behaviors of healthcare providers–using simple language, reducing jargon use, asking questions across a variety of biomedical and social topics–led us to identify a unifying concept in the communication literature. This concept, which encompasses all of these communication behaviors, is known as *message production*, or the process by which a communicator transforms a feeling or thought into a message to share it with other people. While the framework provided by message production did not come directly from the Kirkpatrick Model, we were able to identify it through our focus on patient adherence and holistic interviewing. This identification of message production then allowed us to determine a number of behavioral measures relevant to patient adherence and holistic interviewing. These measures are discussed in further detail in **Section 4.2**.

3.2.3 Learning Objectives

The Kirkpatrick Model includes several components as part of this level: knowledge, skills, attitude, confidence, and commitment (Kirkpatrick and Kirkpatrick, 2016). While all of these components are important to consider when developing training simulations, a focus on skills is likely of most interest for instructional designers, given its potential to overlap with the targeted behaviors from Level 3. Additionally, Kirkpatrick and Kirkpatrick also advise that many of these components can be evaluated simultaneously, so we choose to focus on skill evaluation with the plan to add evaluation of other the components in the future.

Given that we were iterating upon an existing VH training scenario, our first question related to the learning objectives was the degree to which learners were displaying these skills with our VPs currently. We therefore analyzed transcripts from existing interviews using the six metrics we identified from Level 3 Behavior. We analyzed transcripts from real healthcare students to determine if there was any change in these behaviors over time, with the hope that since these learners were enrolled in a clinical practicum course at the time of the interviews, a change in these behaviors would be displayed with the VPs. This analysis is discussed in detail in **Section 4**.

3.2.4 Reactions

The lowest level of the Kirkpatrick Model is the Reactions level. As with many existing training scenarios, we evaluated the Reactions level to some degree before applying the Kirkpatrick Model. This evaluation was done primarily through post-interview survey questions. These questions included items on the medical accuracy of the patient and aspects of the patient the learners found interesting or challenging. We chose to continue this method even after applying the Kirkpatrick Model to keep this level simple, as we felt this best aligned with the Kirkpatrick philosophy to place the higher levels at a greater importance. However, as we continue to develop the VP training scenarios for target patient adherence, questions that explicitly address the communication skills aspect of the training would be helpful. For example, future questions could include asking the learners about the impact of the training on their medical interviewing skills.

4 MATERIALS AND METHODS

After identifying the six message production metrics relevant to patient adherence and holistic interviewing, we then used these metrics to examine retroactively medical communication skills learners' message production behaviors with VPs. We gathered VP interviews from 66 real healthcare students from four previous years to identify any trends in students' message production with VPs over time. The interviews were collected from several cohorts of learners (from the years 2015, 2016, 2018, and 2019) who had had VP interviews integrated into their academic coursework. For each course integration, at the beginning of the semester, students were given an introduction to virtual patient interviewing in VPF2. This introduction covered best practices when using VPF2, including tips such as avoiding the use of pronouns to better match the system's natural language processing or how to track one's progress in an interview. Also at the beginning of the semester, students were asked to complete a background survey with information on their previous experiences interacting with patients and with relevant technology, such as online messaging and videos game use.

After the system introduction and background survey, students began interviewing VPs. The number of VPs interviewed for each course integration, depending on the wishes of the instructors and the goals of the larger studies being conducted, but all students interviewed at least 3 VPs. Of these virtual patients, this work considers the first two interviews, as they are have the most similarities across the different course integrations. Firstly, for all of the cohorts, the first two interviews occurred approximately 1 month apart, and the first and second interviews had no other VP-related tasks between them. A diagram of the course integration tasks represented in this work is provided in **Figure 2**.

The virtual patients interviewed by the students differed, in efforts to coordinate with the instructors what patients they would find most useful to their classes. To investigate whether

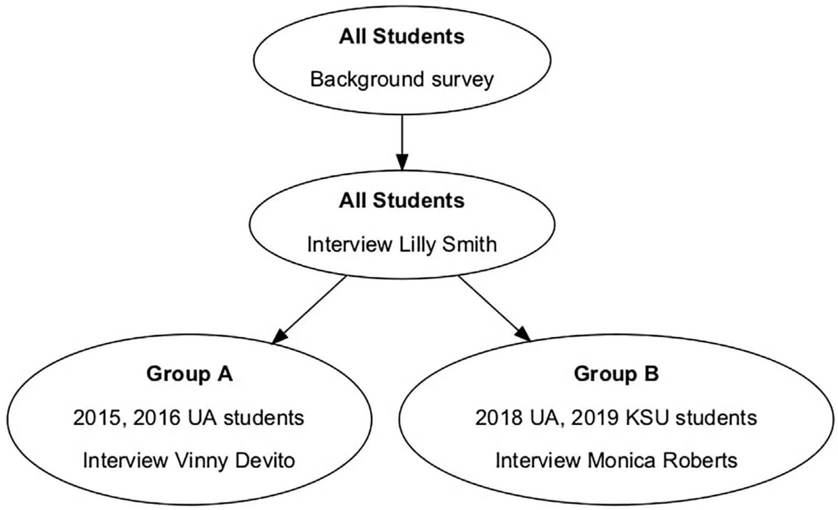

FIGURE 2 | The study tasks completed as part of the Message Production Trends study.

Name	Gender	Age	Diagnosis	Diagnostic difficulty	Interview group
Lilly Smith	Female	65	Parkinson’s disease	4/7 (Neutral)	A and B
Vinny Devito	Male	63	Brainstem stroke	3/7 (Moderately easy)	A
Monica Roberts	Female	38	Head/neck cancer	1/7 (Very easy)	B

TABLE 1 | Demographic information for students in the Message Production Trends study.

Survey item	Group A	Group B	All students
No. of Students	36	30	66
No. of Survey Respondents	36	29	65
Average Age (years)	25.9 ± 4.54	26.5 ± 7.80	26.2 ± 6.16
No. of Female Students	33 female (91.7%)	28 female (96.5%)	61 female (93.8%)
Average Estimate of Patients Interacted With	36 (SD ≈ 17)	27 (SD ≈ 15)	32 (SD ≈ 17)
Received Prior Communication Training	18 No (50.0%)	19 No (65.5%)	34 No (56.6%)

there were any effects on message production due to the students interacting with different virtual patients, students were grouped into two groups based on which patients they interviewed: Group A included the students who interviewed the virtual patients Lilly Smith and Vinny Devito. Group B included the students who interviewed Lilly Smith and Monica Roberts. Group A included students from the University of Auckland in 2015 and 2016, while Group B included students from the University of Auckland in 2018 and Kent State University in 2019.

4.1 Population

Information from the background survey for students in each interview group and across all students is provided in **Table 1**. For this work, 66 students completed the first two virtual patient interviews, but only 65 students completed the background survey. The survey data for the 65 respondents is reported here.

Across both groups, students’ average age was 26.9 ± 6.16 years. The majority of the students in both groups were female, (93.8%). This gender distribution is consistent with real-world speech language pathologists (ASHA, 2020). Students reported interacting with an average of approximately 32 patients with a standard deviation of 17 patients, and a slight majority reported no previous communication skills training (56.9%).

4.2 Metrics for Message Production in Virtual Human Scenarios

Using the framework provided by message production, we identified six measures relevant to patient adherence and holistic interviewing. These six metrics correspond to message production behaviors healthcare students should exhibit when

TABLE 2 | Examples of ICF codes, Flesch Reading Ease scores, and medical words identified for measures in the Message Production Trends study.

Student utterance	ICF code	Flesch reading ease	Medical words identified
do you get a dry mouth	b5104 salivation	116	
describe the sensation during swallowing	b51058 swallowing, other specified	15.6	
do you work	d850 remunerative employment	119	
How about physical activity?	d5701 managing diet and fitness	−8.73	physical, activity
can you feed yourself	e340 personal care providers	97.0	
Are you having difficulty swallowing your medication	e1101 drugs	6.36	medication

interacting with patients. According to communication research, message production has three different categories of assessment: 1) goal attainment, 2) efficiency, and 3) social appropriateness. In the following subsections, we discuss the metrics identified for each category.

4.2.1 Goal Attainment

Key to message production is the role of language as a tool to achieve a goal. Humans do not engage in language use or social interaction as ends themselves but do so to accomplish a goal, such as building rapport (Berger, 2003). Thus, because message production is a goal-driven activity, the degree to which a speaker achieves his or her goal is an important measure. As discussed in **Section 3.2**, an important outcome for our stakeholder and a potential contributor to patient non-adherence was ensuring learners pursue a holistic view of their patients. So, one way we should measure learners' goal attainment is by assessing whether their message production behavior promotes a holistic view of the patient. The measure we identified for this category of message production assessment is the number of unique ICF codes.

A message production behavior that fits this criterion is asking questions on both medical and social topics, and further consultation with our stakeholder introduced us to a systematic set of labels for classifying medical information. This set was the World Health Organization's International Classification of Functioning, Disability, and Health (ICF). The WHO ICF is a framework used to "describe and measure health and disability" (Üstün et al., 2003). As part of this framework, the ICF includes a coding scheme to support a common vocabulary of health topics across disciplines and languages.

Our stakeholder identified a subset of codes from the ICF that related to dysphagia. This subset included a total of ninety-four ICF codes across different categories within the ICF, such as Body Functions and Structures, Environmental Factors, and Personal Factors. Using the subset, we may tag every question asked by learners to determine their coverage of different health topics. Examples of the ICF codes used can be found in **Table 2**. For each learner, the total number of unique ICF codes used in each interview was normalized by ninety-four, the total number of ICF codes being reviewed. A larger number of ICF codes used in a single interview likely indicates a more holistic view of the VP was pursued, as a wider range of topics would have been covered.

4.2.2 Efficiency

The second category of message production assessment is efficiency; speakers can potentially enact multiple strategies to achieve their communication goals, but these strategies may vary in the amount of time and effort needed to enact them (Berger, 2003). To measure efficiency, we identified two metrics: questions per discovery and median question latency.

The questions per discovery metric originates from previous virtual patient literature (Halan et al., 2018) and is the ratio of the number of questions asked by the learner in the interview to the number of discoveries uncovered by the student. This metric reveals how efficiently a learner can uncover the important information in a virtual patient interview. Higher values for questions per discovery indicate less efficient interviewing, as the student had to ask a greater number of questions to uncover discoveries.

Our second efficiency metric, median question latency, is an adaptation of speech latency, a measure that has been used in existing communication literature (Greene and Geddes, 1993). Median question latency is measured in VP interviews by measuring the time interval in seconds between each of a learner's questions to the VP and then taking the median of these intervals. While median question latency is inspired by speech latency in the communication literature, it must be noted that median question latency cannot be compared directly to speech latency, as median question latency contains the additional confound of typing time. Since learners must type their questions to the virtual patient in VPF2, examining the time between each question will also include the time needed to type each question.

4.2.3 Social Appropriateness

The final category of message production assessment is social appropriateness. In general contexts, examples of social appropriateness may include producing messages with the appropriate level of politeness, but as discussed previously in **Section 3.2**, the ability of a healthcare provider to adapt his or her language to promote patient adherence is also important. The two suggestions often given to providers to communicate in a manner that promotes patient adherence are 1) to speak in simple language (Graham and Brookey, 2008; Green et al., 2014; Speer, 2015) and 2) to use less medical jargon (Graham and Brookey, 2008; Oates and Paasche-Orlow, 2009; Green et al., 2014).

To target simple language, we propose two measures: 1) the percentage of learner utterances below the standard reading ease and 2) the percentage of student utterances similar to the virtual patient's. Both of these measures use the Flesch Reading Ease formula (FRE). The FRE has been used in past research to evaluate patient-targeted documents (Williamson and Martin, 2010; Agarwal et al., 2013) and oral health advice (Bradshaw et al., 1975). The FRE uses a text's words per sentence and syllables per word to calculate an overall score (Flesch, 1948). As the score increases, text difficulty decreases. Scores ranging from 60 to 70 are considered "standard" and correspond to an American eighth or ninth grade reading level (Flesch, 1949). The percentage of learner utterances below the standard reading ease (Percent Below Standard) addresses the general difficulty of a learner's utterances by calculating the percentage of utterances that scored below 60, the lower end of the standard range of the FRE.

The percentage of learner utterances similar to the virtual patient's (Percent Similar) was used to measure learner' language difficulty in relation to the virtual patient's. While general recommendations are to use simple language in medical communication, oversimplifying may also be problematic; for example, younger health care providers have been shown to engage in elderspeak with elderly patients (Kemper, 1994). Elderspeak involves changes in lexical complexity, speaking rate, and number of other factors of one's communication and has been associated with inverse health outcomes of the elderly patients it is used with (Williams et al., 2009). Thus, while simple language is important, health care providers should also adapt accordingly to the patient they are currently interacting with. To measure learner adaptability, the reading ease of learners' utterances were compared to the mean of the virtual patient's reading ease. If a learner's utterance was within one standard deviation of the virtual patient's mean reading ease, this utterance was considered "similar" to the virtual patient's. For each learner, the number of similar utterances was normalized by the count of all the learner's utterances to calculate the final metric.

For our final metric, we used the percentage of medical words used by learners to target learners' use of medical jargon. First, a medical word list was created to label the students' transcripts. The medical word list included an 819-word long list created by Lei and Liu in efforts to create an updated academic medical word list (Lei and Liu, 2016). This list was augmented by words pulled from hospital glossaries focused on speech language pathology to ensure coverage of dysphagia-related terms (Cincinnati Childrens, 2021). To determine whether a student's word was a medical word, student utterances were tagged with part-of-speech information using the Python NLTK library (Bird et al., 2009). Since the medical word list only contained nouns, adjectives, and adverbs, the student utterances were filtered down to words of these three parts-of-speech. The remaining words were lemmatized using the lemmatizer provided in the Python NLTK library and then compared against words of the same part-of-speech in the medical word list. For each student, the number of words that matched the medical word list was divided by the total number of words used by the student to produce the final measure.

5 RESULTS

Transcripts from the VP interviews were downloaded from the VPF2 application for processing. While learners may have interacted with each VP multiple times, data was only pulled from a learner's longest transcript to compute the six metrics to prevent artificial inflation of the metrics. For example, when calculating a learner's unique ICF codes, using all of a learner's transcripts may decrease this metric artificially, as the total number of utterances by a learner has no upper limit.

To identify any changes in message production behavior, we ran a mixed-design ANOVA on the six interview metrics. The within-subjects factor was VP interview (Interview 1 and Interview 2) and the between-subjects factor was interview group (Group A or Group B). VP interview was the main effect of interest in this analysis, as any significant effects of VP interview would indicate that there was a change in learners' message production from Interview 1 to Interview 2. Such a finding would suggest that the VP interviews were able to elicit changes in learners' message production. The between-subjects factor of interview group was included to determine if there were any group differences. While group differences were not the main focus of this analysis, we included the between-subjects factor because students came from different institutions and interviewed different virtual patients during their second interview.

Outliers were reviewed for each metric visually using boxplots. Any outliers and their treatment are noted below. Normality, homogeneity of variances, and homogeneity of covariances were assessed by Shapiro-Wilk test, Levene's test of homogeneity of variances, and Box's M test. Any instances in which these assumptions were not met are noted below. A summary of the ANOVA results is provided in **Table 3**.

5.1 Goal Attainment

The assumption of normality was not met for the unique ICF codes metric for the second virtual patient interview, $p < 0.05$, but the mixed-design ANOVA was still performed, as ANOVAs have been shown to be robust to violations to normality (Blanca et al., 2017). There was no significant interaction effect of virtual patient interview and interview group for the unique ICF codes used, $F(1, 64) = 0.506$, $p = 0.479$, partial $\eta^2 = 0.008$. There was not a significant main effect of interview group, $F(1, 64) = 1.78$, $p = 0.18$, partial $\eta^2 = 0.027$, but there was a significant main effect of virtual patient interview, $F(1, 64) = 5.32$, $p = 0.024$, partial $\eta^2 = 0.077$. For students in both interview groups, the percent of unique ICF codes increased from Interview 1 ($19.1 \pm 8.84\%$) to Interview 2 ($21.2 \pm 6.45\%$). The means and standard deviations of the unique ICF codes used in Interview 1 and 2 by both interview groups are available in **Figure 3**.

5.2 Efficiency

For the questions per discovery metric, there was one extreme outlier, as identified by inspection of the SPSS version 26 boxplot. However, exclusion of this outlier did not change the results of the mixed-design ANOVA, so results including this point are presented here. Normality was violated, $p < 0.05$,

TABLE 3 | A summary of the ANOVA results (interaction and main effects) for the Message Production Trends study.

Measure	Interaction effect	VP interview	Interview group
Unique ICF	not significant	Int 1 < Int 2	not significant
Questions per Discovery	not significant	Int 1 > Int 2	not significant
Median Question Latency	not significant	Int 1 > Int 2	A < B
Percent Below	not significant	Int. 1 > Int. 2	A < B
Percent Similar	significant	A: n.s. B: Int 1 < Int 2	Int 1: A > B Int 2: A < B
Percent Med Words	Significant	A: n.s. B: Int 1 < Int 2	Int 1: n.s. Int 2: A < B

Interview groups are abbreviated A and B for Group A, who interviewed Lilly Smithand Vinny Devito, and for Group B, who interviewed Lilly Smith and Monica Roberts
Interviews are abbreviated "Int."
Non-significant results are indicated by "n.s."

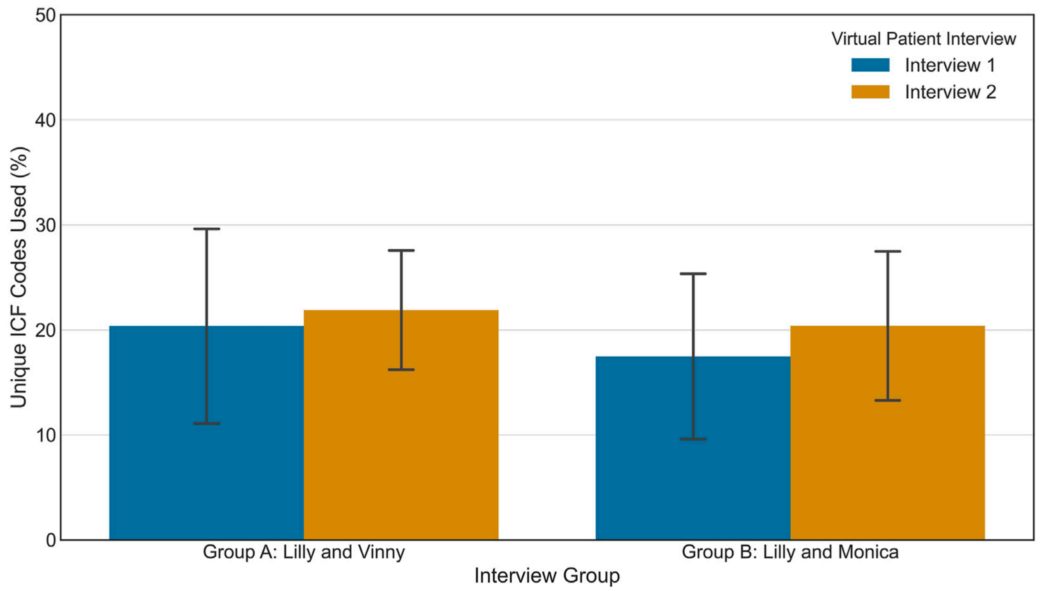

FIGURE 3 | The means and standard deviations of unique ICF codes used for students in the Message Production Trends study.

but the mixed-design ANOVA was still performed. There was no significant interaction effect of virtual patient interview and interview group on questions per discovery, $F(1, 64) = 2.91$, $p = 0.093$, partial $\eta^2 = 0.043$. There was neither a significant main effect of the interview groups, $F(1, 64) = 1.25$, $p = 0.267$, partial $\eta^2 = 0.0190$, but there was a significant main effect of virtual patient interview, $F(1, 64) = 35.7$, $p < 0.005$, partial $\eta^2 = 0.358$. Question per discovery decreased significantly from Interview 1 (5.60 ± 2.71) to Interview 2 (3.65 ± 1.58). The means and standard deviations of questions per discovery for each interview group for Interview 1 and Interview 2 are shown in **Figure 4**.

For the median question latency, shown in **Figure 5**, there was one extreme outlier as identified by inspection of the SPSS version 26 boxplot. Unlike the previous metric, however, inclusion of this outlier did affect the significance results of the interaction effect of mixed-design ANOVA. Analysis reported here therefore excludes the participant with the outlying value, user BK19_08, a member of Group B.

Normality was violated, $p < 0.05$, but the mixed-design ANOVA was still run. There was no significant interaction effect of interview group and virtual patient interview, $F(1, 63) = 3.621$, $p = 0.0616$, partial $\eta^2 = 0.0543$. There was, however, a significant effect of virtual patient interview, $F(1, 63) = 51.5$, $p < 0.005$, partial $\eta^2 = 0.450$. Median question latency significantly decreased from Interview 1 (27.8 ± 11.2 s) to Interview 2 (20.5 ± 6.47 s). Similarly, there was also a significant main effect of interview group, $F(1, 63) = 10.5$, $p = 0.002$, partial $\eta^2 = 0.143$. Group B had a significantly higher median question latency (27.5 ± 10.0 s) than Group A (21.4 ± 8.81 s).

5.3 Social Appropriateness

For percent of learner utterances below the standard reading ease (Percent Below), the assumption of normality was not met for students in Group B during Interview 2, $p < 0.05$. The mixed design ANOVA was still performed. There was no significant interaction effect of virtual patient interview and interview group on Percent Below, $F(1, 64) = 0.203$, $p = 0.654$, partial $\eta^2 = 0.003$.

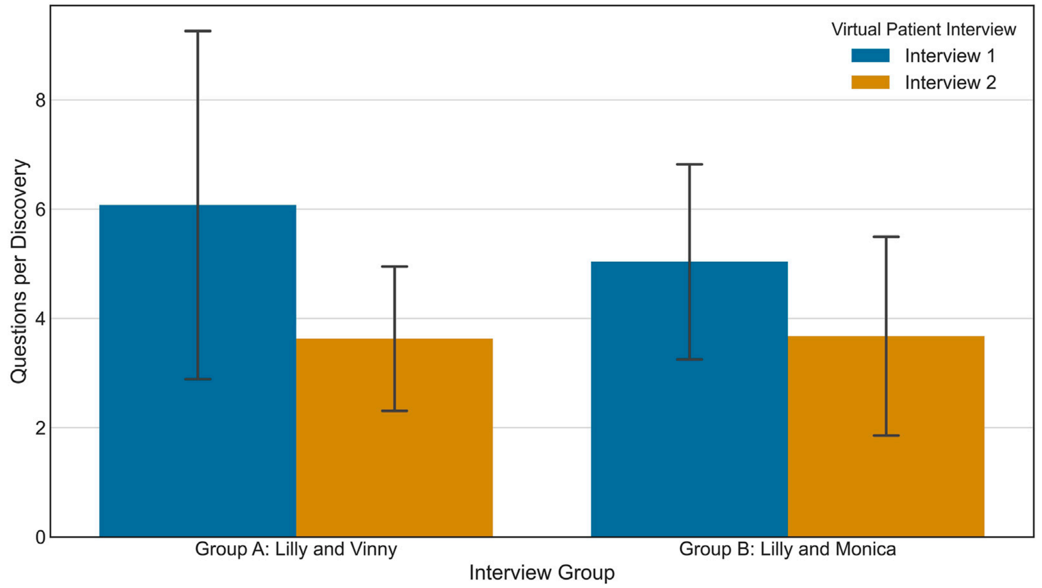

FIGURE 4 | The means and standard deviations of questions per discovery for students in the Message Production Trends study.

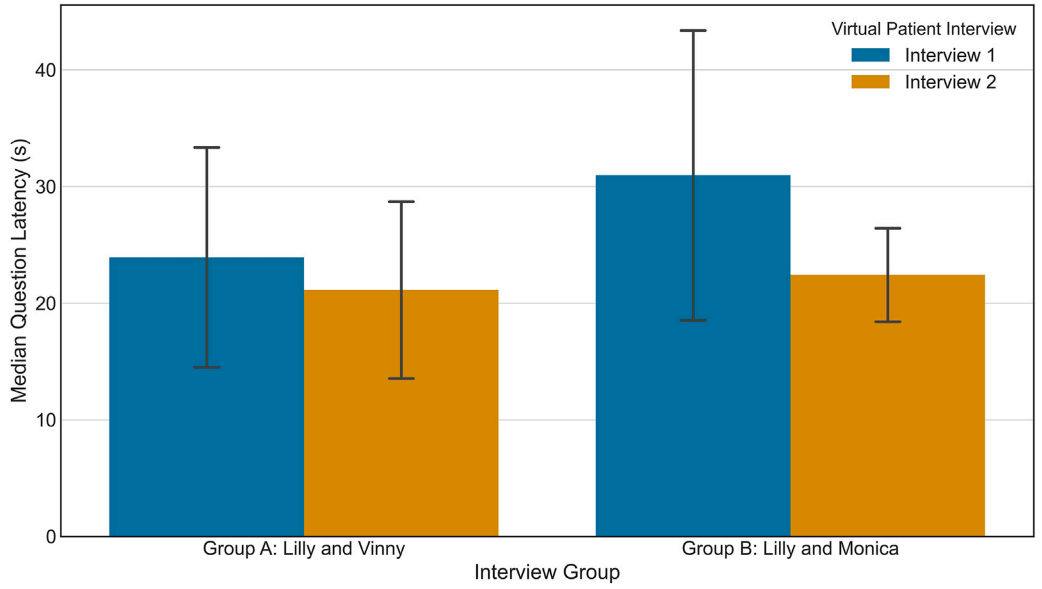

FIGURE 5 | The means and standard deviations of median question latency for students in the Message Production Trends study.

There was a significant main effect of virtual patient interview, $F(1, 64) = 5.30$, $p < 0.025$, partial $\eta^2 = 0.076$. For students in both groups, the average of Percent Below for Interview 1, 22.8 ± 8.78%, was significantly higher than the average for Interview 2, 20.3 ± 7.67%. There was also a significant main effect of interview group, $F(1, 64) = 15.4$, $p < 0.005$, partial $\eta^2 = 0.194$. Averaged across both interviews, Group B's Percent Below measure was significantly greater than Group A's. This trend may be observed in **Figure 6**.

For percent of learner utterances similar to the virtual patient's (Percent Similar), the assumption of normality was not met for students in Group a during Interview 1, $p < 0.05$. The mixed design ANOVA was still performed. There was a significant interaction effect of virtual patient interview and interview group on Percent Similar, $F(1, 64) = 27.7$, $p < 0.005$, partial $\eta^2 = 0.302$.

Follow-up analysis for the main effect of interview group revealed that there was a significant difference in Percent

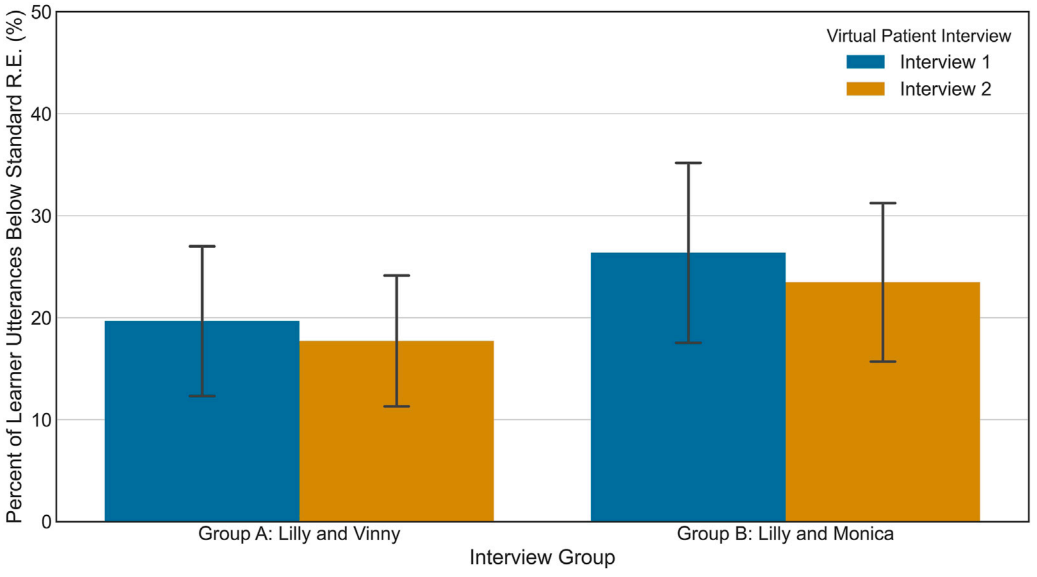

FIGURE 6 | The means and standard deviations of percent of learner utterances below standard reading ease for students in the Message Production Trends study.

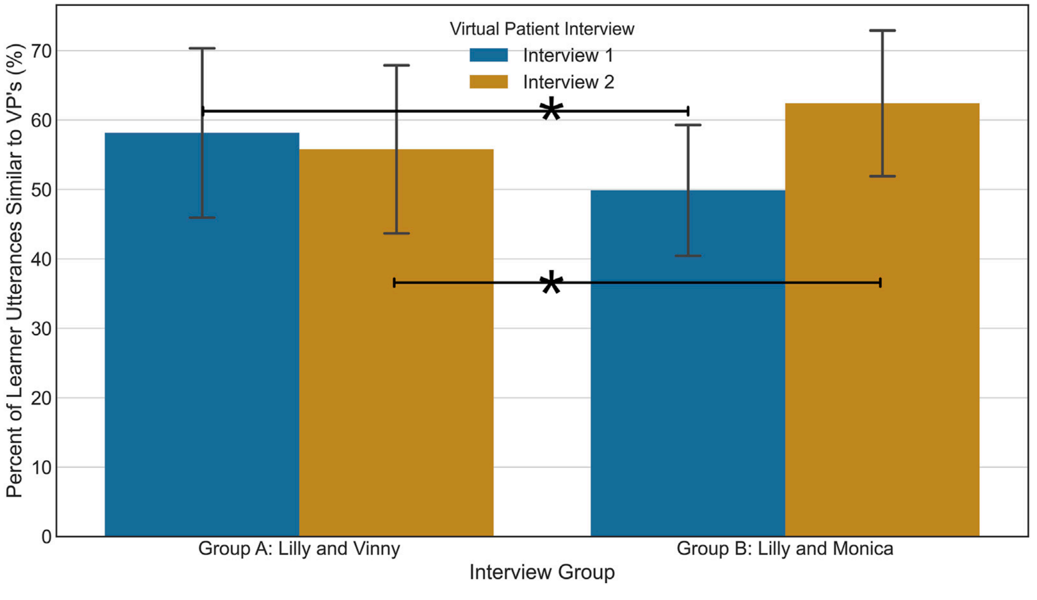

FIGURE 7 | The means and standard deviations of percent of learner utterances similar to the virtual patient's for students in the Message Production Trends study.

Similar between the groups during Interview 1, F(1, 64) = 8.952, $p = 0.004$, partial $\eta^2 = 0.123$. For Interview 1, Group A's Percent Similar measure was significantly greater (58.2 ± 12.4%) than Group B's (49.9 ± 9.59%). A significant difference was also at present at Interview 2, F(1, 64) = 5.36, $p = 0.024$, partial $\eta^2 = 0.077$, but in the opposite direction. Group B's Percent Similar metric (62.4 ± 10.7%) was significantly greater than Group A's (55.8 ± 12.3%).

Follow-up analysis for the main effect of virtual patient interview shows that only Group B displayed a significant change in Percent Similar over the two interview, F(1, 29) = 30.2, $p < 0.005$, partial $\eta^2 = 0.510$. Group B's Percent Similar measures increased from an average of 49.9 ± 9.59% in Interview 1 to 62.4 ± 10.7% in Interview 2. There was no significant change for Group A over Interview 1 and Interview 2, F(1, 35) = 1.82, $p = 0.186$, partial $\eta^2 = 0.049$. These trends may be observed in **Figure 7**.

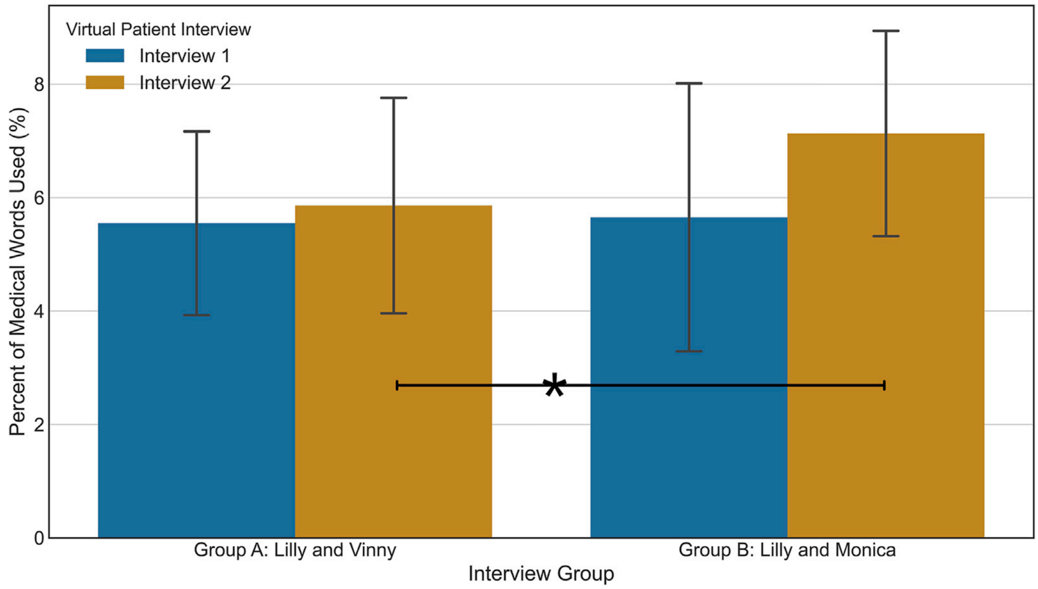

FIGURE 8 | The means and standard deviations of percent of medical words used for students in the Message Production Trends study.

Finally for the percent of medical words used (Percent Medical), shown in **Figure 8**, the assumptions of normality and homogeneity of covariances were not met, $p < 0.05$, but the mixed design ANOVA was still performed. There was a significant interaction effect of virtual patient interview and interview group on the percent of medical words used, $F(1, 64) = 4.13$, $p = 0.046$, partial $\eta^2 = 0.061$.

Follow-up analysis for the main effect of interview group reveals there was a significant difference between the two groups at Interview 2, $F(1, 64) = 7.40$, $p = 0.008$, partial $\eta^2 = 0.104$. Group B's average for percent of medical words used was significantly higher ($7.13 \pm 1.84\%$) than Group A ($5.86 \pm 1.93\%$). This difference was not present in Interview 1, $F(1, 64) = 0.045$, $p = 0.833$, partial $\eta^2 = 0.001$.

Follow-up analysis for the main effect of virtual patient interview revealed only a significant change for Group B, $F(1, 29) = 7.79$, $p = 0.009$, partial $\eta^2 = 0.212$. The percent of medical words used by Group A did not change significantly over the course of the interviews, $F(1, 35) = 1.23$, $p = 0.276$, partial $\eta^2 = 0.034$.

6 DISCUSSION

Our discussion of our results is broken into two subsections **Section 6.1**, discusses potential trends in learners' message production and how they relate to known patterns in message production, while **Section 6.2** discusses what the results of this work suggest for the application of the Kirkpatrick Model to other learning scenarios based in virtual environments.

6.1 Discussion of Learners' Message Production With Virtual Patients

Students' message production in the VP interviews demonstrated changes in some measures, as there were several significant main effects of virtual patient interview. The main effects of virtual patient interview indicated that students' goal attainment and efficiency metrics changed significantly from Interview 1 to Interview 2. A main effect of virtual patient interview was also found for the Percent Below metric, one of the social appropriateness measures. These changes in students' message production suggest that students ask questions on more topics, ask these questions more efficiently, and use less complicated language in Interview 2 than Interview 1. Based on these findings, virtual human interviews elicited changes in a variety of message production behaviors and may be useful in measuring students' message production behavior throughout a semester.

Interestingly, for both the median question latency and the Percent Below metric, in addition to significant effect of virtual patient interview, there was also a significant difference between interview groups. As stated previously, a between-subjects factor was included in this analysis because students were required to interview different virtual patients in Interview 2 and because students came from different academic institutions. Group B (Lilly and Monica) included some students from Kent State University in the United States while Group A (Lilly and Vinny) only contained students from the University of Auckland in New Zealand. Cultural or environmental differences may have prompted some of the Group B students to produce messages in a manner different than those in Group A. However, further analysis with more students from different institutions would be needed to investigate this properly, as the majority of the students in this analysis came from the same institution, the University of Auckland.

The results for the social appropriateness metrics revealed additional differences in message production. For the remaining two metrics–Percent Similar and Percent Med Words–there were significant interaction effects. For both metrics, Group B experienced a significant increase from Interview 1 to Interview 2. At Interview 2, Group A's values are also significantly less than Group B's. In contrast to the Percent Below measure, in which we saw an overall difference in Group B (Lilly and Monica) compared to Group A (Lilly and Vinny), the influencing factor here seems to be isolated to Interview 2, suggesting that the changes in these metrics may be due to speaking to a different VP.

One potential reason that students in Group B spoke in a more similar language complexity to the VP during the second interview may be due to the ages and genders of the virtual patients interviewed. Previous research in linguistics shows that speakers "align" their speaking more closely to their speaking partners' if the partner is considered an "in-group" member (Unger, 2010). In other words, in conversation, one speaker may mimic another speaker more if the second speaker is perceived to be similar. This perception of in-group versus out-group may have been present when students interviewed the virtual patients. Lilly Smith (Interview 1) is depicted as a 65 year-old female, while Monica Roberts (Group B, Interview 2) is depicted as a 38 year-old female. Since Monica Roberts is closer in age to the participants, the participants may have tried to match Monica more than Lilly in terms of language complexity. Such a perception may have affected the percentage of medical words used as well.

6.2 Overall Discussion

Using the metrics identified from the Kirkpatrick Model related to holistic interviewing and patient adherence, we demonstrated that the VPs elicited changes in students' message production behavior over time. From this finding, we identify two contributions. Firstly, our work adds to the existing ability of VHs to elicit real-world behaviors from participants, as demonstrated in the works discussed previously (Slater et al., 2006; Cassell et al., 2009). Secondly, based on our application of the Kirkpatrick Model to identify how these behavioral measures were made, we find support for our suggestion to introduce educational VH simulations at the Behavioral level in the Kirkpatrick Model.

The ability to include educational VH simulations at later stages in the Kirkpatrick Model could have a great impact for developers of these applications. Because the metrics derived using the Kirkpatrick Model originate from important objectives in the educational context itself, this process provides some assurance that the measures are meaningful to what is being learned. Further, by incorporating behavioral measures into the VH scenario, there is the potential to lessen the gap between the Learning and Behavior levels in the Model, as learners will be able to engage in the critical behaviors while still interacting with the VH itself. The VH-based training may also be used as a Behavior level monitoring solution, which is critical to ensure trainees continue to apply training in real world settings. While future work will be needed to evaluate the general ability of virtual environments to blend aspects of the Learning and Behavior levels, our work provides initial support for this line of inquiry.

7 CONCLUSION

In this work, we recommend the use of the Kirkpatrick Model as a framework to evaluate educational and training applications using virtual environments. Specifically, we investigated the use of non-fully immersive, conversational VHs to evaluate learner behavior change during training by using the Kirkpatrick Model to identify behavioral measures that can be evaluated both in virtual environments and in the real-world. By incorporating behavioral measures into our VP-based desktop application, we hope to lessen the potential gap between the virtual simulations and the behaviors learners should perform in the real world. Our work provides a new perspective on measuring behavior as compared to the standard Kirkpatrick Model, which advises that learners' behaviors may only be observed while in real-world scenarios.

In our application of the Kirkpatrick Model, we derived six metrics related to healthcare students' real-world behaviors (Level 3) that promote holistic interviewing and patient adherence (Level 4 Results). These six metrics were then used to evaluate healthcare students' message production with VPs over the course of an academic semester. We found significant changes in three of the six metrics. While follow-up research would be needed to confirm that these changes reflect students' message production trends with real patients, we view this finding as encouraging: the behavior metrics motivated by the Kirkpatrick Model have some sensitivity to students' language behavior and can be also be reused to evaluate students' language behavior with real patients later on. Additional work will be needed to validate this approach, but we find support for our new perspective of the Kirkpatrick Model to observe behavior level measures with non-fully immersive VH technology. Additional work is needed to further validate our approach in fully-immersive simulations. Future work can also investigate the effects of measuring behavior level measures in simulation by comparing learner behaviors across virtual environments and reality, as well by tracking larger metrics such as those found in the Model's results level.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institutional Review Board, University of Florida and The University of Auckland Human Participants Ethics Committee (UAHPEC 016700). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SC conducted the user studies and carried out the data analysis. BL and AM contributed to the design of the user studies. AM was the instructor of record for the University of Auckland students and the stakeholder in the Kirkpatrick Model process. All authors contributed to the authoring and conceptualization of the manuscript.

ACKNOWLEDGMENTS

The authors thank Heng Yao for his role in running the user study and data collection in 2018, as well as the Virtual Experiences Research Group for their feedback and expertise in both VPF2 development and manuscript feedback. Finally, the authors thank Dr. Ali Barikroo for his guidance in integrating the virtual patient interviews into his course.

REFERENCES

Agarwal, N., Hansberry, D. R., Sabourin, V., Tomei, K. L., and Prestigiacomo, C. J. (2013). A Comparative Analysis of the Quality of Patient Education Materials from Medical Specialties. *JAMA Intern. Med.* 173, 1257–1259. doi:10.1001/jamainternmed.2013.6060

Alaraj, A., Lemole, M. G., Finkle, J. H., Yudkowsky, R., Wallace, A., Luciano, C., et al. (2011). Virtual Reality Training in Neurosurgery: Review of Current Status and Future Applications. *Surg. Neurol. Int.* 2, 52. doi:10.4103/2152-7806.80117

Alliger, G. M., Tannenbaum, S. I., Bennett, W., Jr, Traver, H., and Shotland, A. (1997). A Meta-Analysis of the Relations Among Training Criteria. *Personnel Psychol.* 50, 341–358. doi:10.1111/j.1744-6570.1997.tb00911.x

Asha, A. S.-L.-H. A. (2020). *Profile of ASHA Members and Affiliates, Year-End 2019*. Tech. rep. Available at: https://www.asha.org/siteassets/surveys/2001-2021-member-and-affiliate-profile-trends.pdf.

Bassi, L., Benson, G., and Cheney, S. (1996). *Trends: Position Yourself for the Future*. Alexandria, VA: American Society for Training and Development.

Beal, M. D., Kinnear, J., Anderson, C. R., Martin, T. D., Wamboldt, R., and Hooper, L. (2017). The Effectiveness of Medical Simulation in Teaching Medical Students Critical Care Medicine. *Sim Healthc.* 12, 104–116. doi:10.1097/SIH.0000000000000189

Berger, C. R. (2003). "Message Production Skill in Social Interaction," in *Handbook of Communication and Social Interaction Skills* (Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers).

Bird, S., Klein, E., and Loper, E. (2009). *Natural Language Processing with Python*(Beijing:Cambridge [Mass.]: O'Reilly)OCLC. 1st ed edn. ocn301885973.

Blanca, M. J., Alarcón, R., Arnau, J., Bono, R., and Bendayan, R. (2017). Non-normal Data: Is ANOVA Still a Valid Option? *Psicothema* 29, 552–557. doi:10.7334/psicothema2016.383

Bradshaw, P. W., Ley, P., Kincey, J. A., and Bradshaw, J. (1975). Recall of Medical Advice: Comprehensibility and Specificity. *Br. J. Soc. Clin. Psychol.* 14, 55–62. doi:10.1111/j.2044-8260.1975.tb00149.x

Brogden, H. E., and Taylor, E. K. (1950). The Theory and Classification of Criterion Bias. *Educ. Psychol. Meas.* 10, 159–183. doi:10.1177/001316445001000201

Cassell, J., Geraghty, K., Gonzalez, B., and Borland, J. (2009). Modeling Culturally Authentic Style Shifting with Virtual Peers."in Proceedings of the 2009 international conference on Multimodal interface (Cambridge, MA: ICMI-MLMI '09) 135. doi:10.1145/1647314

Cincinnati Childrens (2021). *Speech-Language Pathology Glossary*.

Cohen, D., Sevdalis, N., Taylor, D., Kerr, K., Heys, M., Willett, K., et al. (2013). Emergency Preparedness in the 21st century: Training and Preparation Modules in Virtual Environments. *Resuscitation* 84, 78–84. doi:10.1016/j.resuscitation.2012.05.014

Delisle, M., Ward, M. A. R., Pradarelli, J. C., Panda, N., Howard, J. D., and Hannenberg, A. A. (2019). Comparing the Learning Effectiveness of Healthcare Simulation in the Observer versus Active Role: Systematic Review and Meta-Analysis. *Sim Healthc.* 14, 318–332. doi:10.1097/SIH.0000000000000377

Dikeman, K. J., and Riquelme, L. F. (2002). Food for Thought. *Perspect. Swal Swal Dis. (Dysph)* 11, 31–35. doi:10.1044/sasd11.3.31

DiMatteo, M. R. (2004). Evidence-based Strategies to foster Adherence and Improve Patient Outcomes. *JAAPA* 17, 18

Dukes, L. C., Pence, T. B., Hodges, L. F., Meehan, N., and Johnson, A. (2013). "Sidnie," in *Proceedings of the 2013 International Conference on Intelligent User Interfaces* (New York, NY, USA: Association for Computing Machinery)), 395–406. doi:10.1145/2449396.2449447

Flesch, R. (1948). A New Readability Yardstick. *J. Appl. Psychol.* 32, 221–233. doi:10.1037/h0057532

Flesch, R. F. (1949). *Art of Readable Writing Publisher*. Harper.

Foster, A., Chaudhary, N., Kim, T., Waller, J. L., Wong, J., Borish, M., et al. (2016). Using Virtual Patients to Teach Empathy. *Sim Healthc.* 11, 181–189. doi:10.1097/sih.0000000000000142

Fox, J., Arena, D., and Bailenson, J. N. (2009). Virtual Reality. *J. Media Psychol.* 21, 95–113. doi:10.1027/1864-1105.21.3.95

Gordon, D. (1996). MDs' Failure to Use plain Language Can lead to the Courtroom. *CMAJ* 155, 1152

Grabowski, A., and Jankowski, J. (2015)., 72. Publisher: Elsevier, 310–314. doi:10.1016/j.ssci.2014.09.017Virtual Reality-Based Pilot Training for Underground Coal Miners*Saf. Sci.*

Graham, S., and Brookey, J. (2008). Do Patients Understand? *Perm J.* 12, 67–69. doi:10.7812/tpp/07-144

Green, J. A., Gonzaga, A. M., Cohen, E. D., and Spagnoletti, C. L. (2014). Addressing Health Literacy through clear Health Communication: A Training Program for Internal Medicine Residents. *Patient Educ. Couns.* 95, 76–82. doi:10.1016/j.pec.2014.01.004

Greene, J. O., and Geddes, D. (1993). An Action Assembly Perspective on Social Skill. *Commun. Theor.* 3, 26–49. doi:10.1111/j.1468-2885.1993.tb00054.x

Halan, S., Sia, I., Crary, M., and Lok, B. (2015). "Exploring the Effects of Healthcare Students Creating Virtual Patients for Empathy Training," in *International Conference on Intelligent Virtual Agents* (Springer), 239–249. doi:10.1007/978-3-319-21996-7_24

Halan, S., Sia, I., Miles, A., Crary, M., and Lok, B. (2018). "Engineering Social Agent Creation into an Opportunity for Interviewing and Interpersonal Skills Training," in *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems* (Stockholm, Sweden).

Kemper, S. (1994). Elderspeak: Speech Accommodations to Older Adults. *Aging Neuropsychol. Cogn.* 1, 17–28. doi:10.1080/09289919408251447

Kidd, L. I., Knisley, S. J., and Morgan, K. I. (2012). Effectiveness of a Second Life Simulation as a Teaching Strategy for Undergraduate Mental Health Nursing Students. *J. Psychosoc Nurs. Ment. Health Serv.* 50, 28–37. doi:10.3928/02793695-20120605-04

Kirkpatrick, D. L., and Craig, R. (1970). *Evaluation of Training*. Evaluation Of Short-Term Training in Rehabilitation , *35Publisher*. New York: ERIC.

Kirkpatrick, D. L. (1954). *Evaluating Human Relations Programs for Industrial Foremen and Supervisors*. Madison, Wisconsin: Doctoral, University of Wisconsin–Madison.

Kirkpatrick, J. D., and Kirkpatrick, W. K. (2016). *Kirkpatrick's Four Levels of Training Evaluation*. Alexandria, VA: Association for Talent Development.

Kleinsmith, A., Rivera-Gutierrez, D., Finney, G., Cendan, J., and Lok, B. (2015). Understanding Empathy Training with Virtual Patients. *Comput. Hum. Behav.* 52, 151–158. doi:10.1016/j.chb.2015.05.033

Kundhal, P. S., and Grantcharov, T. P. (2009). Psychomotor Performance Measured in a Virtual Environment Correlates with Technical Skills in the Operating Room. *Surg. Endosc.* 23, 645–649. doi:10.1007/s00464-008-0043-5

Lei, L., and Liu, D. (2016). A New Medical Academic Word List: A Corpus-Based Study with Enhanced Methodology. *J. English Acad. Purposes* 22, 42–53. doi:10.1016/j.jeap.2016.01.008

Li, A., Montaño, Z., Chen, V. J., and Gold, J. I. (2011). Virtual Reality and Pain Management: Current Trends and Future Directions. *Pain Manag.* 1, 147–157. doi:10.2217/pmt.10.15

Loukas, C., Nikiteas, N., Kanakis, M., and Georgiou, E. (2011). Evaluating the Effectiveness of Virtual Reality Simulation Training in Intravenous Cannulation. *Simulation Healthc.* 6, 213–217. doi:10.1097/sih.0b013e31821d08a9

Martin, L. R., Williams, S. L., Haskard, K. B., and DiMatteo, M. R. (2005). The challenge of Patient Adherence. *Ther. Clin. Risk Manag.* 1, 189
Oates, D. J., and Paasche-Orlow, M. K. (2009). Health Literacy. *Circulation* 119, 1049–1051. doi:10.1161/CIRCULATIONAHA.108.818468
Raij, A., Johnsen, K., Dickerson, R., Lok, B., Cohen, M., Bernard, T., et al. (2006). Interpersonal Scenarios: Virtual\\approx Real?" in IEEE Virtual Reality Conference (VR 2006). IEEE, 59–66.
Rossen, B. H. (2011). *Design and Evaluation of Conversational Modeling Methods for Interpersonal Simulation*. Ph.D. thesis, Gainesville, FL: University of Florida.
Schmidt, B., and Stewart, S. (2009). Implementing the Virtual Reality Learning Environment. *Nurse Educator* 34, 152–155. doi:10.1097/nne.0b013e3181aabbe8
Shaw, A., Ibrahim, S., Reid, F., Ussher, M., and Rowlands, G. (2009). Patients' Perspectives of the Doctor-Patient Relationship and Information Giving across a Range of Literacy Levels. *Patient Educ. Couns.* 75, 114–120. doi:10.1016/j.pec.2008.09.026
Slater, M., Pertaub, D.-P., Barker, C., and Clark, D. M. (2006). An Experimental Study on Fear of Public Speaking Using a Virtual Environment. *CyberPsychology Behav.* 9, 627–633. doi:10.1089/cpb.2006.9.627
Speer, M. (2015). "Using Communication to Improve Patient Adherence," in *Communicating with Pediatric Patients and Their Families: The Texas Children's Hospital Guide for Physicians, Nurses and Other Healthcare Professionals* (Houston, USA: Texas Children's Hospital), 221–227.
Suárez, G., Jung, S., and Lindeman, R. W. (2021). Evaluating Virtual Human Role-Players for the Practice and Development of Leadership Skills. *Front. Virtual Real.* 2, 31. doi:10.3389/frvir.2021.658561
Unger, L. (2010). *The Social Role of Linguistic Alignment with In-Group and Out-Group Members Publisher*. Edinburgh: The University of Edinburgh.
Ustün, T. B., Chatterji, S., Bickenbach, J., Kostanjsek, N., and Schneider, M. (2003). The International Classification of Functioning, Disability and Health: a New Tool for Understanding Disability and Health. *Disabil. Rehabil.* 25, 565–571. doi:10.1080/0963828031000137063
Van Wyk, E., and De Villiers, R. (2009). Virtual Reality Training Applications for the Mining Industry. *Proc. 6th Int. Conf. Comput. graphics, virtual reality, visualisation interaction Africa*, 53–63. doi:10.1145/1503454.1503465
Waisman, Y., Siegal, N., Chemo, M., Siegal, G., Amir, L., Blachar, Y., et al. (2003). Do Parents Understand Emergency Department Discharge Instructions? A Survey Analysis. *Isr. Med. Assoc. J.* 5, 567
Williams, K. N., Herman, R., Gajewski, B., and Wilson, K. (2009). Elderspeak Communication: Impact on Dementia Care. *Am. J. Alzheimers Dis. Other Demen.* 24, 11–20. doi:10.1177/1533317508318472
Williamson, J. M. L., and Martin, A. G. (2010). Analysis of Patient Information Leaflets provided by a District General Hospital by the Flesch and Flesch-Kincaid Method. *Int. J. Clin. Pract.* 64, 1824–1831. doi:10.1111/j.1742-1241.2010.02408.x
Wouda, J. C., and van de Wiel, H. B. M. (2013). How to Attain Expertise in Clinical Communication? *Paediatric Respir. Rev.* 14, 213–218. doi:10.1016/j.prrv.2013.04.005
Wouda, J. C., and van de Wiel, H. B. M. (2012). The Communication Competency of Medical Students, Residents and Consultants. *Patient Educ. Couns.* 86, 57–62. doi:10.1016/j.pec.2011.03.011
Wouda, J. C., Zandbelt, L. C., Smets, E. M. A., and van de Wiel, H. B. M. (2011). Assessment of Physician Competency in Patient Education: Reliability and Validity of a Model-Based Instrument. *Patient Educ. Couns.* 85, 92–98. doi:10.1016/j.pec.2010.09.007
Xie, B., Liu, H., Alghofaili, R., Zhang, Y., Jiang, Y., Lobo, F. D., et al. (2021). A Review on Virtual Reality Skill Training Applications. *Front. Virtual Real.* 2, 49. doi:10.3389/frvir.2021.645153
Zaveri, P. P., Davis, A. B., O'Connell, K. J., Willner, E., Schinasi, D. A. A., and Ottolini, M. (2016). Virtual Reality for Pediatric Sedation: A Randomized Controlled Trial Using Simulation. in *Cureus* (San Francisco, CA: Publisher: Cureus Inc)

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Virtual Reality

HYPOTHESIS AND THEORY
published: 28 April 2022
doi: 10.3389/frvir.2022.694048

I Know It Is Not Real (And That Matters) Media Awareness vs. Presence in a Parallel Processing Account of the VR Experience

Tilo Hartmann[1]* *and Matthias Hofer*[2]

[1]*Department of Communication Science, Vrije Universiteit Amsterdam (VU), Amsterdam, Netherlands,* [2]*Faculty of Arts and Social Sciences, Department of Communication and Media Research, University of Zurich, Zurich, Switzerland*

Inspired by the widely recognized idea that in VR/XR, not only presence but also encountered plausibility is relevant (Slater, Phil. Trans. R. Soc. B, 2009, 364 (1535), 3549–3557), we propose a general psychological parallel processing account to explain users' VR and XR experience. The model adopts a broad psychological view by building on interdisciplinary literature on the dualistic nature of perceiving and experiencing (mediated) representations. It proposes that perceptual sensations like presence are paralleled by users' belief that "this is not really happening," which we refer to as *media awareness*. We review the developmental underpinnings of basic media awareness, and argue that it is triggered in users' conscious exposure to VR/XR. During exposure, the salience of media awareness can vary dynamically due to factors like encountered sensory and semantic (in)consistencies. Our account sketches media awareness and presence as two parallel processes that together define a situation as a media exposure situation. We also review potential joint effects on subsequent psychological and behavioral responses that characterize the user experience in VR/XR. We conclude the article with a programmatic outlook on testable assumptions and open questions for future research.

Keywords: media awareness, presence, perceptual sensation, representation, pictorial competence, parallel processing, virtual reality

Edited by:
Richard Skarbez,
La Trobe University, Australia

Reviewed by:
Sam Srauy,
Oakland University, United States
Ekaterina Prasolova-Førland,
Norwegian University of Science and Technology, Norway

***Correspondence:**
Tilo Hartmann
t.hartmann@vu.nl

Specialty section:
This article was submitted to Virtual Reality and Human Behaviour, a section of the journal Frontiers in Virtual Reality

Received: *12 April 2021*
Accepted: *18 March 2022*
Published: *28 April 2022*

Citation:
Hartmann T and Hofer M (2022) I Know It Is Not Real (And That Matters) Media Awareness vs. Presence in a Parallel Processing Account of the VR Experience. Front. Virtual Real. 3:694048. doi: 10.3389/frvir.2022.694048

INTRODUCTION

When scholars explicate user's experience of Virtual Reality (in the following we simply speak of VR or the VR experience, but we believe that our ideas extend to any extended reality/XR technology and experience), they traditionally focus on the sensation of presence. In this manuscript, when we talk about *presence,* we follow the definition provided by Lee (2004), who defines presence as "psychological state in which virtual (para-authentic or artificial) objects are experienced as actual objects in either sensory or nonsensory ways" (p. 37). For the sake of simplicity, we focus on presence as a unitary concept that includes subtypes such as spatial, social, and self-presence (Lee, 2004). Accordingly, presence entails users' sensation of owning a virtual body, and of "being there" in a virtual or virtually augmented space, perhaps with social others feeling co-present. In general, we regard presence as a highly automatic, cognitively non-taxing, mostly sensory-driven perceptual sensation or feeling that is introspectively accessible (Schubert, 2009).

While presence has been highlighted as a defining part of the VR experience, scholars in the field also frequently noted that users of VR still stay at least partially aware of the mediated nature of their experience (i.e., they know that "this is not real or really happening"; e.g., ISPR, 2001). What is this *media awareness*, as we call it in the present article, and when or how does it shape the VR experience? Do two users who feel equally present, but differ in their media awareness, have a different overall user experience? In the present paper, we address these important questions. Our central proposition is that a comprehensive conceptualization of the VR experience (and, potentially, even the experience of any mediated representation or content) must emphasize both users' perceptual sensations like presence and their media awareness, and recognize how both jointly shape users' overall experience.

We are not the first scholars suggesting this idea. In fact, in a widely influential and recognized article, Slater (2009) proposed that users' responses to VR can only be fully understood if not only presence is taken into account, but also users' perceived plausibility. According to Slater, users respond to VR as if it was real only if they both feel spatially present ("place illusion," PI) and if they simultaneously feel that events in the scenario refer to their presence, respectively that events are actually taking place ("plausibility illusion," Psi). Slater concluded his conceptualization with a call for further research on users' perceived plausibility: The "area of Psi is now a more fruitful and challenging research area than PI" (p. 3555). In the present article, we try to answer Slater's call by re-positioning plausibility and presence in a more general parallel processing account of users' VR experience, which is inspired by existing research on the dualistic nature of representations (e.g., Grodal, 2002; Nieding et al., 2017), and converges with recent discussions by other VR scholars (e.g., Gonzalez-Franco and Lanier, 2017; de Gelder et al., 2018; Pan and Hamilton, 2018).

We proceed in five steps to develop a new theoretical look on the VR experience. First, we briefly review Slater's influential conceptualization of the plausibility (vs. place) illusion and the revision developed by Skarbez (2016). Second, because we suggest considering plausibility as part of a bigger picture, we broaden the view (beyond plausibility, and beyond VR) by reviewing existing interdisciplinary research on the dualistic nature of users' experience of mediated representations. This research suggests that users' experience derives from their perceptual sensations or intuitive feelings and their higher-order beliefs or knowledge about what is happening. In a third step, we explicate media awareness as users' belief that "this is not really happening", illustrate its developmental underpinnings, and discuss how it is cued at the onset and during media exposure. Fourth, we discuss how media awareness and perceptual sensations like presence might be related to each other. Consequently, we explicate how both might jointly affect the overall user experience. We conclude the article in a fifth step by looking at how the proposed framework can guide and inspire future research.

The Plausibility Illusion

Presence is the hallmark of the VR experience, also in comparison to other media channels that only evoke this sensation to a lesser degree, if at all. Yet, VR users' experience is not fully or adequately described by only focusing on presence (Pan and Hamilton, 2018). This fact has been most prominently addressed to date in a widely recognized article by Slater (2009). In this article, Slater focuses on the question when or why users respond realistically to VR. This is a relevant question, because VR is often said to trigger life-like experiences and it is increasingly used as a tool to train or study real-world behavior (Fox et al., 2009). According to Slater (2009), users respond realistically to VR if they experience place illusion and plausibility illusion, which he considers as two "orthogonal components" (p. 3549). The place illusion is a perceptual illusion that refers to "the sense of being there" (commonly addressed as spatial, physical, or tele-presence). This factor has received a lot of attention in the past and is by now relatively well understood (see for overviews, e.g., Haans and IJsselsteijn, 2012; Hartmann et al., 2015; Gonzalez-Franco and Lanier, 2017).

In contrast to presence, the plausibility illusion received much less scholarly attention and is less well understood to date. According to Slater (2009), this illusion refers to users' sensation "that the scenario being depicted is actually occurring" (p. 3549), even if users know for sure that this is not true. According to Slater (2009), the plausibility illusion results from the extent the virtual environment "acknowledges" users' presence in the world (i.e., shadows cast by a user's avatar, or an agent's eye-gazing towards the avatar, etc.). Furthermore, the illusion results from "the overall credibility of the scenario being depicted in comparison with (users') expectations" (p. 3549). Hence, the plausibility illusion is "concerned with the "reality" of the situation depicted" (p. 3556), which implies that users' expectations are supported. Yet, Slater also recognizes that even if the scenario appears highly realistic and plausible, "at a higher cognitive level (users) know that nothing is "really" happening, and they can consciously decide to modify their automatic behaviour accordingly" (p. 3554).

Slater's (2009) approach provides an intriguing elaboration of the VR experience. Yet, while the approach offers a lot of valuable insights, some questions remain about exact concept definitions and their integration into existing literature. For example, Slater's (2009) approach focuses on when or why users respond realistically to VR (i.e., as if they were in a non-mediated situation). A bit confusingly perhaps, responding realistically is addressed as presence (Skarbez et al., 2017), while what was (and probably still is) commonly understood as one important type of presence (i.e., the "feeling of being there") is dubbed the place illusion. Furthermore, while the idea of the plausibility illusion is very intriguing, its exact operationalization stayed perhaps a bit tentative in the original approach (see also Skarbez et al., 2017). The way it is introduced, the plausibility illusion seems potentially overlapping with the outcome (i.e., users responding as if the scenario was real, Berthiaume et al., 2021). Furthermore, the plausibility illusion seems to be closely related to perceived realism, a multi-dimensional concept that is well established in the literature (Popova, 2010). In addition, the fact that users always stay aware at a higher cognitive level that "this is not really happening" is noted yet not fully elaborated in Slater's original approach, and remains somewhat disconnected to the other ideas, e.g., about plausibility.

In our view, some of this ambiguity surrounding the plausibility illusion has been resolved by Skarbez and colleagues (Skarbez, 2016; Skarbez et al., 2017, see also; Gilbert, 2016). The authors propose that the plausibility illusion builds on coherence (i.e., the extent to which a virtual scenario "behaves reasonably" or consistent to users' expectations). If the VR technology frequently fails to support users' expectations, it is unlikely that they will respond to the virtual environment as if it was real. Coherence thus appears to be central in understanding the user experience, yet it also remains an ambiguous concept. For example, Skarbez et al. (2017) focus on coherence as a system factor, while others consider it a user factor (Berthiaume et al., 2021). For instance, in perceived realism research (Hall, 2003), coherence has been considered as users' perceived external (consistency with real-life knowledge) and internal (consistency within the description) plausibility of the media depiction (Busselle and Bilandzic, 2008). While Skarbez et al. (2017) consider coherence as a factor that is specifically affecting the plausibility illusion, other scholars regard coherence as central to the general user experience, including potential presence experiences (Seth et al., 2012; Latoschik and Wienrich, 2021). How coherence affects users' VR experience apparently is a topic of debate, but by highlighting coherence as a central factor, Skarbez and colleagues helped both refining Slater's original idea and integrating it more firmly into existing research.

The present account aims to contribute and further expand these attempts to explicate the VR experience. We believe that present theorizing in this area can benefit from broadening the theoretical view and incorporating insights from a wider range of existing literature (e.g., about how users perceive and respond to media representations in general). We believe that such a broader approach moves the focus away from plausibility onto *media awareness*, a concept that we introduce in the present article. According to our notion, whenever encountering mediated content, users simultaneously feel that "this is real or happening" while knowing that their experience is mediated. If adapted to VR, we propose that "feeling that this is real" refers to users' sensation of presence, while "knowing this is not real" refers to users' media awareness. We are convinced that we can only reach a comprehensive understanding of users' VR experience if we model how both users' presence sensations and their media awareness jointly shape the overall experience. We propose that plausibility, in turn, matters as a determinant of media awareness.[1]

REVIEWING THEORETICAL ACCOUNTS OF THE DUALISTIC MEDIA EXPERIENCE

Our general approach is inspired by a central idea expressed in various conceptualizations of how users experience mediated representations. These mostly disconnected approaches stem from interdisciplinary research strands like film or book studies in the humanities, but also research on art perception, optical illusions, philosophy, and research from (perceptual, media, cognitive, developmental) psychology. Most of these approaches target the experience of specific mediated representations, like sketched figures, photos, film, narratives (e.g., in books) or VR, while some (e.g., Wolf, 2017) set out to model the experience of any (mediated) representation. As diverse as they might be, a core idea expressed in all of these approaches is that the user experience is inherently *dualistic*. In the following, we review a couple of related relevant concepts or approaches in more detail:

- A first relevant related concept is the *aesthetic illusion*, which is mostly studied in media (film/text) studies in the humanities and arts (see for an overview see for instance Wolf, 2014; Koblížek, 2017). Wolf (2017) defines the aesthetic illusion as primarily "a *feeling*, with variable intensity, of being imaginatively and emotionally immersed in a represented world and of experiencing this world as a presence (. . .) in an as-if mode, that is, in a way similar (but not identical) to real life. At the same time, however, this impression of immersion is counterbalanced by a latent rational distance resulting from a (. . .) (*media-*)*awareness* of the difference between representation and reality" (p. 32, italics added). The quote reveals what scholars on the aesthetic illusion define as the central dualistic nature of the media experience, namely users' intuitive sensations of an apparent reality and their parallel awareness of the mediated nature of their intuitive sensations.
- In the realm of picture perception, "seeing-in" (Wollheim, 1998) represents another relevant concept addressing the dualistic or parallel nature of the mediated experience. Wollheim argues that when looking at a picture, viewers have two simultaneous experiences: they are aware of the represented object (e.g., a house) and the way the object is represented (e.g., red oil paint). Wollheim claims that the two experiences are not independent, but two aspects of a single experience which he refers to as *two-foldedness*. According to Nanay (2005), "(a) visual experience of an agent is 'twofold' if she is simultaneously aware of both the represented object and the medium of representation" (p. 263).
- Relatedly, psychological-developmental research on symbolic or *pictorial competence* highlights the dualistic nature of representations (or symbolic artifacts). "Every symbolic artifact is an object in and of itself, and at the same time it also stands for something other than itself" (DeLoache et al., 2003, p. 114). According to DeLoache et al. (2003), to "understand and use a symbol, dual representation is necessary—one must mentally represent both facets of the symbol's dual reality, both its concrete characteristics and its abstract relation to what it stands for" (p. 114).
- In his general theory of film perception and visual aesthetic (the *PECMA Flow*), rooted in cognitive film studies, Grodal (2002); Grodal (2006) regards cinematic experiences not as processing of representations, but as primary (real-world)

[1]We agree with Latoschik and Wienrich (2021) that plausibility is also a general determinant of presence. However, because the present approach introduces and focuses on media awareness, we do not comprehensively discuss determinants of presence in the present paper.

experiences "although we know that this seeing is induced by artificial means" (2006, p. 3). According to Grodal (2002) the more salient users' media awareness, the more it "is added to, and enriches, the phenomenal experience" (p. 72). However, according to the PECMA flow, recalling that the processed depictions are not real requires some cognitive effort.

- Communication science scholars argued that media users can switch either between an involved *reception mode* (accepting the presented world; thinking "within" the logic of the media offering) or analytical reception mode (reflecting about the media offering, (Michelle, 2007; Suckfüll and Scharkow, 2009; Frey, 2018). The latter, sometimes also referred to as psychical or aesthetic distance (Cupchik, 2001), includes considering how a certain film or scene was produced (Suckfüll and Scharkow, 2009). Relatedly, Frey (2018) distinguishes an *experiential mode* of reception from a *thinking or non-experiential* mode of reception. The thinking mode is characterized by mental effort, and it can result in either greater belief and acceptance or in greater disbelief and rejection of the "apparent reality" (p. 500) suggested by the media depiction.
- In philosophy, Gendler (2008); Gendler (2019) proposed to distinguish alief from belief. According to Gendler, alief is an automatic or habitual belief-like attitude. "Charles believes that he is sitting safely in a chair in a theater in front of a movie screen (but) the alief has roughly the following content: 'Dangerous two-eyed creature heading towards me! H-e-l-p . . . ! Activate fight or flight adrenaline now!'" (Gendler, 2008, p. 637). As the example shows, for Gendler alief represents users' acceptance of the depiction, whereas belief represents their co-existing knowledge that "this is just mediated".
- Many scholars also already stressed the dualistic nature of users' *VR experience.* However, rather than providing a full account of media awareness, past literature often referred to VR users' "knowing that this is not real" as a curious side aspect. Recently, however, several scholars started focusing more closely on the dualistic nature of the VR experience. Gonzalez-Franco and Lanier (2017), for example, discuss users' "partial awareness of (the presence) illusion" (p. 5). They speculate that high plausibility and strong sensory saturation provided by VR, and high cognitive load among users might reduce media awareness. Similarly, de Gelder et al. (2018), p. 2) argue that in VR users are "in a state (...) where knowledge of the unreality of the VR world and belief in its experiential reality coexist". de Gelder et al. (2018), p. 2) stress that users' knowledge of the unreality of VR "denotes the special cognitive or epistemological status of the VR experience". Other scholars, too, noted this dualistic nature of the VR experience. We identify related ideas, for example, in Turner's (2016) argument that experiencing presence requires pretense, and in Waterworth and Tjostheim's (2021) argument that VR users believe what is happening is real, "except in the sense that at some level, (they) know the virtual reality is a simulation" (p. 23).

Our literature review does not claim to be comprehensive. Yet, it shows how scholars from largely disconnected fields converge on a strikingly similar idea about how users process and experience (mediated) representations, and hence also VR. The idea is that users' experience of representations is inherently dualistic: Users intuitively process, perceive, and experience represented content "as if it was real or unmediated," while simultaneously, and in varying intensity, staying cognitively aware that their experience is triggered by a representation. Accordingly, we also assume that VR users might automatically feel present, while simultaneously staying aware that this sensation is triggered by VR technology.

PRESENCE

In the present approach we endorse a broad conceptualization of presence that entails various forms such as users' feeling spatially present in VR or users perceiving artificial agents to be physically co-present in their augmented real environment. In general, presence is a conscious perceptual sensation or feeling (one feels present, something feels present). Presence builds on the interplay of external sensory stimulation of the VR/XR system, and users' motor actions, respectively the internal interoceptive and proprioceptive signals they accompany (e.g., see "sensorimotor contingencies," Slater, 2009, p. 3549). In the logic of the predictive coding paradigm, presence arises if the predictions about external and internal sensory signals that accompany motor action are so accurately matched by technology that any residual error can be "explained away" by the brain (Seth et al., 2012).

As a perceptual sensation, we think presence is an inevitable user response to any correctly calibrated VR system. Neither pretense nor a related suspension of disbelief (Wirth et al., 2007; Waterworth and Tjostheim, 2021) might be necessary to foster presence. For example, a proper VR system will always make human users feel spatially present. To provide another example, eye-gazing of an artificial agent that is augmented into users' actual environment will inevitably trigger a subtle feeling of social or co-presence (Senju and Johnson, 2009). However, our central argument is that in the context of VR exposure, these automatic perceptual presence sensations are always accompanied by the belief, or awareness, that they are triggered by human-made technology. Users' media awareness provides the (cognitive) backdrop based on which perceptual sensations like presence are interpreted.

WHAT IS MEDIA AWARENESS?

In short, media awareness is about users' salient belief that "this is not really happening" during VR (or any media) exposure. We define media awareness as the salience of users' propositional belief or conviction that their experiences during exposure are based on human-made technology. In other words, if being media aware, users believe and are conscious of the fact that what they perceive and experience in the present situation is largely determined by human-made technology rather than by authentic (non-artificial) stimuli that were actually present here and now. More specifically, media awareness can imply different things. It can imply that users believe that they currently do not perceive an object or event directly, but through a medium or interface (like in live events, tele-surgery, or when navigating drones). In addition, it can imply that users believe that currently encountered objects or events are not actually existing (in space and time) but are non-authentic or fictional. These rather stable and firm higher-order *cognitive* beliefs, which we deem central to media awareness, can be distinguished from lower-order *perceptual* beliefs that might originate from perceptual sensations like presence and that only have a tentative status (e.g., see similarly Gilbert et al., 1993; Grodal, 2002; Kahneman and Frederick, 2002; Gawronski and Bodenhausen, 2014; Herschbach, 2015; Gendler, 2019).

If media awareness is about the belief that "this is not really happening," when and how is it activated? And to what extent does it stay in mind during exposure? We claim that media awareness is 1) activated when individuals consciously initiate a media exposure episode and that it, subsequently, stays minimally salient in mind; we refer to this as *basic awareness*. Furthermore, we claim that 2) media awareness can dynamically vary in salience up-and-above this basic level, which we refer to *dynamically salient media awareness* We start by discussing basic media awareness and its developmental underpinnings.

Basic Media Awareness

Whenever users consciously approach a medium or representation (and recognize it as such, e.g., a photo, a VR headset, or a hologram) the belief that "this is not really happening" will be accessed from propositional knowledge and will be activated. We assume that the belief is held in working memory (i.e., "the ensemble of components of the mind that hold a limited amount of information temporarily in a heightened state of availability for use in ongoing information processing," Cowan, 2017, p. 1163). Once activated, the belief stays in mind during media exposure (e.g., through attentional refreshing, Camos et al., 2018), and thus establishes a permanent baseline level of media awareness. We think basic media awareness is what scholars mean if they say that users, despite feeling present, still *know* that they are using technology (ISPR, 2001; Slater et al., 2006; Slater, 2009).

We assume that basic media awareness has important consequences. For example, if individuals are looking at the hologram of a duck, they perhaps will establish a tentative perceptual sensation that "there is a duck" (Zeimbekis, 2015). They might even walk around the hologram-duck and look at it from different angles. Yet, the belief that "this is not really happening" will embed this sensation, thus shaping the overall experience of their response towards (the representation of) the duck. Accordingly, basic media awareness should affect the construction of meaning, and how a situation is subjectively interpreted. It subjectively defines the overall situation as a media exposure situation, and provides a cognitive backdrop based on which perceptual sensations are interpreted. To use the words of Grodal (2002, p. 72), we think that through basic media awareness users' "knowledge of "reference" is added to, and enriches, the phenomenal experience".[2]

Developmental Underpinnings of Media Awareness

Being "media aware" requires competence and learning. This competence is acquired during child development. Perceiving representations, from a simple Necker cube to a moving 3D-object in VR, is easy and effortless, particularly if they sufficiently match authentic objects in appearance and functionality. The representation's sensory information (e.g., visual depth cues) feeds into quickly activated, hard-wired or heavily ingrained perceptual mechanisms that immediately create the perceptual sensation. Therefore, the represented object often springs to mind easily, and a vivid perception of it unfolds naturally. We effortlessly and automatically perceive a cube as a 3D-object or see a face when looking at a picture. In fact, as Zeimbekis (2015) notes in the context of picture perception, in most cases, the representation provides the natural and the medium the non-natural perception (Grodal, 2002; Wolf, 2017).

Accordingly, it is perhaps not surprising that we as human beings first have to learn to become aware of mediated representations (i.e., to recognize them, to understand what they imply, and how to use them, Flavell et al., 1983; Schlottmann, 2001). This ability is addressed in the literature as pictorial or symbolic competence (DeLoache et al., 2010). Humans routinely start developing symbolic competence very early in life and continue to develop this skill, as they grow older. Infants, for example, if exposed to objects in pictures, first tend to try to grasp these objects. They fail to accurately distinguish the representation from its authentic counterpart, as they have not yet learned what a picture is. Once they understand that objects on pictures are "these things that look like the actual object but can't be grasped" (Grodal, 2002), young children learn two important things. First, these sensory objects are called a photo. From thereon, they can categorize and interpret their photo-induced sensations adequately. That is, they start developing a theory or conceptualization of media. This resembles other important developments taking place at this age, such as the development of a theory of mind (Flavell, 2000). Second, they learn that the accurate response to a photo is to point to objects rather than trying to grasp them (DeLoache et al., 2010). Of course, from thereon symbolic competence will continue to extend.[3] In summary, we consider symbolic competence as the central ability that develops throughout ontogenesis to allow an individual becoming "media aware". It is the developmental underpinning of users'

[2]We speculate that this will also be the case in the future, in which we perhaps encounter perfect virtual simulations. Just as awareness about dreaming distinguishes lucid dreaming from non-lucid dreaming (Quaglia and Holecek, 2018), basic media awareness should still distinguish exposure to perfect virtual simulations from exposure to non-mediated reality.

[3]Consider, for example, evolving literacy about how to respond to "true-depth" pictures in 3D-movies or VR. It is striking to observe that even adults first try to grasp objects, before realizing that their hands move through objects and these are not truly tangible (Flavell et al., 1983; Ross, 2015).

propositional knowledge or belief that "this is not really happening".

Dynamically Salient Media Awareness

During media exposure, levels of media awareness can also dynamically fluctuate. It appears that the belief that "this is not really happening" can recede to the back of mind or stay on top of mind (e.g., Jacobs and Silvanto, 2015), while never dropping below a certain baseline level. In other words, users might sometimes be very aware that "this is just mediated" and sometimes barely aware, while never forgetting that "this is just mediated". We assume that these dynamic shifts of media awareness might be driven by user factors, like users' motivated attention allocation, for example if they actively want to recall that this is not real (Busselle and Bilandzic, 2008). However, they might also arise from the interplay of the medium and user. For example, due to inconsistencies or flaws noted by perceptual processes (Gilbert, 2016; Skarbez et al., 2017), users might intuitively sense that something is wrong or odd, or unreal. In the search for an explanation for this sensation of unrealness users are likely to recruit basic media awareness (i.e., the firm propositional belief that experienced perceptual sensations originate from media technology). In other words, something seems strange, but this irritation can be smoothly explained by the already activated belief that "this is not really happening". As a consequence, perceptual sensations of unrealness might shift basic media awareness back into the focus of attention.[4]

Triggering Dynamically Salient Media Awareness During Exposure

Developing symbolic competence, and acquiring related propositional knowledge about media representations, arises from encounters that individuals have with stimuli in their environment that seem somehow different to their authentic counterparts. Individuals encounter inconsistencies (i.e., violations of their expectations that are grounded in experience of the authentic world). Hence, these inconsistencies require a new classification of the encountered stimuli (see also Gilbert, 2016). Environmental stimuli that regularly trigger inconsistencies are categorized as non-authentic, represented or mediated. Subsequently, the same inconsistencies might cue this category if encountered again, also during media exposure. It still seems to be an open research question which inconsistency cues exactly mark objects, events, or situations as odd, unreal, or mediated. While we are unaware of an overarching psychological account to date, intriguing yet still tentative ideas have been offered in specific contexts, like in picture perception (Zeimbekis, 2015) or film reception (Grodal, 2002). In addition, a couple of systematic, yet still speculative, elaborations exist in the VR context (Lombard and Ditton, 1997; Timmins and Lombard, 2005; Gonzalez-Franco and Lanier, 2017). Closely following these ideas (e.g., Gilbert, 2016; Skarbez, 2016), we assume two clusters of inconsistency cues that plausibly categorize something as odd, unreal or mediated - and thus also affect the salience of media awareness, namely 1) sensory inconsistency (i.e., the extent to which represented objects or events fail to match expectations about their authentic counterparts in terms of sensory information and affordances, and the overall visibility of the medium), and 2) semantic inconsistency (i.e., the extent to which represented objects and events are unexpected or seem unlikely, given the present context or situation). In general, in line with Gonzalez-Franco and Lanier (2017) we propose that these violations of sensory or propositional consistency might increase the salience of media awareness during exposure.

Sensory inconsistency refers to users' sensing of the representation or interface. We distinguish two processes. First, sensory inconsistency can refer to the extent that a depicted entity (e.g., an object) fails to provide the sensorimotor contingencies or affordances that are expected from interaction with its authentic counterpart (see for related ideas predictive coding; Seth et al., 2012; sensory power and consistency, Skarbez, 2016; Gonzalez-Franco and Lanier, 2017; reality status, Grodal, 2006; authenticity, Gilbert, 2016). If expectations are not met and mismatches cannot be easily explained away or integrated (Biocca et al., 2001), the representation reveals itself. For example, Zeimbekis (2015) argues that pictures do not provide "binocular disparity" and thus no stereoscopic depth. Accordingly, they "do not engage the motion-guiding visual system" (dorsal (motion) vs. ventral vision, p. 319) of the brain although the user might see or rather construct depth. "So perhaps the dorsal system dedicated to navigation 'knows' that the picture is a more or less flat object, while at the same time the ventral system picks up the volumetric contents and depth relations from the picture's surface" (Zeimbekis, 2015, p. 320). Hence pictures, or the sensations they evoke, need to be categorized by the user as something different than their authentic counterparts. In terms of VR representations, the situation is very similar, although the user can navigate and therefore the dorsal system is active. However, also in VR representations important cues are missing (e.g., tactile, temperature, or olfactory cues) that a user likely expects from authentic counterparts. These mismatches might cue VR as something mediated. Potentially, the lower the match between the affordances provided by the VR (or any) representation and the expectations based on its authentic counterpart, the more likely it is that media awareness is cued. In addition, technical glitches in a VR (e.g., rendering problems, frozen screens) might violate a number of expectations and thus potentially represent a very strong trigger of media awareness.

[4]We agree with others that also in non-mediated situations individuals can have the intuitive perceptual sensation that "this is not really happening". For example, individuals suffering from derealization disorder can feel detached from the world, and experience it as muted or fake, as if seen through a see-through glass wall (Dokic and Martin, 2017). Similar experiences of unrealness are reported by severely stressed witnesses of disastrous events (Timmins and Lombard, 2005). However, while at the perceptual level sensations of unreality might be similar in mediated vs. non-mediated situations, we assume that the overall experience will be different in terms of physiological, emotional, or behavioral responses. We believe that this difference is due to the fact that in non-mediated situations, these sensations of unrealness cannot be attributed to human-made technology, warranting more concerning explanations (like bodily dysfunction).

Second, building on non-supported affordances, sensory inconsistency refers to the extent that the medium reveals itself to the user, not based on imperfect sensory representation and unsupported affordances, but based on the visibility of the interface itself (e.g., visible canvas or pixels, a TV frame, the edge of VR goggles). The medium (or the interface) inevitably reveals itself, we believe, if the user consciously initiates the exposure situation. However, the medium might also reveal itself during exposure. For example, users might shift their attention to the cover of a book or frame of a TV, and thus feel reminded of the mediated origin of their experience. In VR, the user seems to be more enveloped by the technology (field of view covered by headset, unrestricted movements, headphones; Slater and Wilbur, 1997), thus potentially lowering the visibility of the medium (Lombard and Ditton, 1997). However, the weight of the headset, tangible cables, and visible pixels are among the cues that potentially reveal the medium during VR exposure, too. In addition, cross-cutting sensory information from the non-mediated environment (e.g., hearing a shout, bumping into an object) or from the body (cybersickness, Rebenitsch and Owen, 2016) might trigger media awareness, if users fail to successfully integrate this information and instead shift their attention onto the interface in their attempt to make sense of the situation.

The second cluster of factors that might dynamically increase media awareness—*semantic inconsistency*—represents a more cognitive cluster than the sensory-based first cluster. We believe that media awareness might also depend on the extent to which depicted entities or events fail to meet users' expectations that they derive from their propositional knowledge. Hence, media awareness might vary based on how plausible or likely users find encountered objects or events (see "plausibility" or "coherence," Gonzalez-Franco and Lanier, 2017; Skarbez et al., 2017; Latoschik and Wienrich, 2021). If encountered entities or events are very unexpected, or deemed highly implausible or unlikely, users' sense-making attempts might increase their media awareness. We propose two different types of plausibility judgments, respectively violated expectancies. First, semantic inconsistency can refer to the extent to which encountered objects or events seem (im-) plausible in light of users' real-world knowledge and expectations. This type has been addressed as external plausibility in perceived realism research (Busselle and Bilandzic, 2008; Popova, 2010; Hofer et al., 2020), and, if referring to a social world, as social realism (Lombard and Ditton, 1997). How irritating violations of external plausibility are depends on how much users expect the encountered format or genre to display reality (i.e., to match their real-world propositional knowledge). External plausibility violations (e.g., a flying elephant) might trigger media awareness only if users expect the format or genre to offer high semantic affinity with the real world (e.g., like documentaries or news, Busselle and Bilandzic, 2008). Second, semantic inconsistency can refer to the extent to which encountered objects or events appear (in-) consistent within the logics of the presented story or environment. This type has been addressed as internal plausibility (Popova, 2010) or narrative realism (Busselle and Bilandzic, 2008). For instance, even if the format offers fiction, a flying elephant might appear implausible, if the narrative previously emphasized that elephants cannot fly and users subsequently fail to come up with a compelling reason for the flying elephant.

The provided list of cues that trigger media awareness suggests that virtually any media technology reveals itself once the user consciously activates it (e.g., from opening a book to putting on the VR headset). During exposure, virtually all existing media provide imperfect sensory fidelity and do not support all expected affordances (e.g., impossibility to look behind objects or to touch them). Often, the interface stays visible during exposure. Zeimbekis (2015), p. 321) argues that to date perhaps only an old media technique, namely trompe-l'oeil, provides perfect illusions (or delusions) in which users might be completely unaware. Trompe-l'oeils do not require conscious exposure, and (if the vantage point is right) do not violate sensory and semantic consistency. The question is whether any other media technology, like VR, will be able to delude users one day, so that they are completely unaware of using a medium. Presumably, this would require XR technology like glasses that we would commonly wear, and we would then forget about wearing. This device, which then almost must become a permanent part of one's body, might augment reality perhaps in such a sensory- and semantically consistent way that we might be completely unaware that we encounter non-authentic objects, people, or events (Biocca, 1997).

Sensory and semantic inconsistency suggest that users become more fully "media aware" if the system does not sufficiently support their expectations. However, potentially users might also vary the salience of media awareness completely voluntarily, independent of the content they encounter and its perceived consistency. Accordingly, for very different reasons, users might also be simply motivated to actively recall that "this is not really happening", and thus momentarily refresh media awareness and increase its salience in working memory (Camos et al., 2018). These motivated recalls might be backed up by guiding the attentional focus onto "evidence" that this is not really happening. An example would be a user who feels strongly co-present with a scary monster in VR, and thus experiences strong fear as a response. This user might be motivated to enhance media awareness and actively recall that "this is not really happening".

Reducing Dynamically Salient of Media Awareness

Sensory-based and semantic inconsistency are likely to heighten media awareness during exposure. In addition, users might voluntarily heighten media awareness if they are motivated to recall "that this is not really happening." However, which factors might potentially lower media awareness beyond the mere absence of the above-mentioned factors?

Gonzalez-Franco and Lanier (2017) hypothesize that greater familiarity with VR, and higher cognitive load, might both decrease media awareness. We agree that greater familiarity might decrease media awareness. We find this assumption plausible for two reasons. First, more familiar users might either encounter fewer technical issues or need to pay less

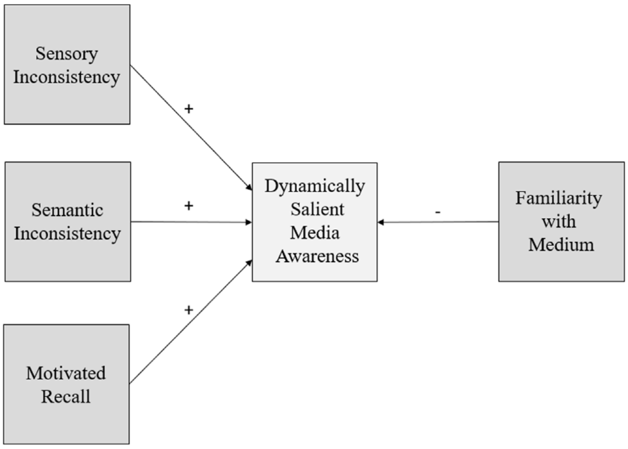

FIGURE 1 | Factors triggering (+) and factors reducing (-) dynamically salient media awareness. On the left, sensory inconsistency and semantic inconsistency are possible features of VR/XR, and motivated recall is a user factor that might enhance the salience of dynamical varying media awareness. On the right, familiarity with the VR/XR is a user factor that likely reduces dynamically salient media awareness.

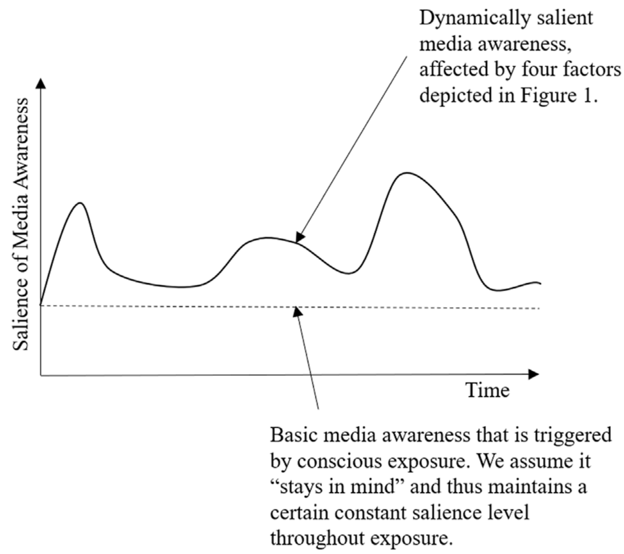

FIGURE 2 | Development of media awareness over time throughout exposure. Conscious media exposure activates a basic constant level of media awareness. During media exposure, several factors (introduced in **Figure 1**) influence the dynamically salient part of media awareness. We assume media awareness never drops below baseline levels, however.

attention to the interface than less familiar users. Second, with repeated exposure users might adapt their expectations to what is commonly displayed in VR, thus encountering fewer surprises (e.g., about missing affordances or semantic inconsistencies, Gonzalez-Franco and Lanier, 2017; Berthiaume et al., 2021). Familiarity might thus plausibly affect media awareness, but future research is necessary to test this assumption.

While we agree with Gonzalez-Franco and Lanier (2017) that cognitive load is also an interesting factor to examine, we are skeptical that staying media aware, or simply recalling that "this is not really happening" qualifies as a cognitively taxing activity. Therefore, we also doubt that cognitive load would impede media awareness. In the absence of empirical evidence, we think it remains speculative, if not doubtful, that the belief that "this is not really happening" becomes less salient, or is less easily refreshed in working memory (Camos et al., 2018), if processing resources are largely occupied by paying attention to the displayed environment and objects in VR.

Intermediate Summary

In summary, our argument is that throughout early ontogenesis we develop the competence to distinguish representation from their authentic counterparts. While achieving this skill, we also learn about the stimuli (interfaces, media technologies) that bring forth related and possibly "strange" sensory experiences, and their names (book, TV, smartphone, VR). Conscious initiation of exposure activates the belief that subsequent perceptual sensations like presence, no matter how compelling, are not really happening (in the sense that they originate in the real world), but can be attributed to the technology. This belief stays in mind as a basic media awareness and allows the user to subjectively interpret the situation as a media exposure situation. Therefore, we think that users never respond to encountered representations the same way as they would do if they believed that "this is really happening" (we will return to this point again later). However, salience of the belief might also vary throughout exposure, heightened by sensory and semantic inconsistencies, and by pro-active or motivated recall, and lowered potentially with greater familiarity (see **Figure 1**). An additional factor that might reduce media awareness is cognitive load, yet we are skeptical that keeping in mind that "this is not really happening" is cognitively taxing, and therefore we did not include this factor in our model depicted in **Figure 1**. The moderating impact of the belief "that this is not really happening" on the overall user experience (discussed below, see also **Figure 2**) might partly depend on how salient it is. **Figure 2** depicts the unfolding of media awareness over the course of a VR exposure episode. The x-axis represents time and the y-axis represents the salience of media awareness.

A PARALLEL PROCESSING ACCOUNT OF THE VR EXPERIENCE: PRESENCE VS. MEDIA AWARENESS

So far, we roughly suggested that media awareness co-occurs with the perceptual sensation of presence, and that they together define the typical VR experience. In the remainder of this article, we elaborate these two ideas. Before we discuss how presence and media awareness jointly shape the overall VR experience, in this section, we refine the idea that both result from two parallel processes during exposure. So far, leaning on literature on perception, we considered presence an outcome of bottom-up perception, and media awareness an outcome of top-down cognition or knowledge. At the same time, we think both presence and media awareness can also be linked to two different processing systems that underlie people's reasoning, judgments, and beliefs, and are prominently discussed in psychological research on dual processing (Evans, 2007). In light of dual processing, we think that presence stems from associative processing and media awareness from propositional processing (Gawronski and Bodenhausen, 2011).

According to this view (Hartmann, 2012; Hofer, 2016; Krcmar and Eden, 2019), presence, as a perceptual sensation, is the result of quick, effortless, and automatic sensory-driven perceptual or so-called System-1 processing in the brain. System-1 processing requires no specific training or literacy; it is an in-born facility human beings share with other animals and that is already commonly utilized by infants. System-1 processing gives rise to an intuition, gut feeling, or tentative perceptual belief (Kahneman, 2012). Schubert (2009) proposes that users' feeling of spatial or social presence resembles such a gut feeling or tentative perceptual belief. In contrast, System-2 operates based on knowledge, and rule-based logical or analytical processing. System-2 has been linked to uniquely human facilities, such as hypothetical thinking, mental simulations, and detection of illusions (Evans and Stanovich, 2013). Accordingly, we propose that media awareness is evoked by System-2 processing.

An important yet thorny question is how both processes interact with each other during exposure. For example, does System-1 processing, and hence presence as an output, interfere with System-2 processing, thus affecting how media aware users are during exposure? To address this question, we must also look at the extent to which both processes co-occur throughout exposure. Interaction between both processes seems only possible when both processes co-occur during exposure, but not if one of the processes is muted. Research on dual processing distinguishes parallel-competitive (Smith and DeCoster, 2000; Sloman, 1996) and default-interventionist (Evans and Stanovich, 2013) dual processing theories. When applied to the present case, we think that presence vs. media awareness might better be modeled as resulting from parallel-competitive than from default-interventionist processing.

Parallel-Competitive Processing

In light of the a parallel-competitive dual processing notion, presence and media awareness would be the outcomes of two processes that constantly co-occur throughout media exposure, yet run largely independent from each other and do not causally affect each other (see also "simultaneous contradictory belief", Sloman, 1996, p. 11; see also for a related discussion of visual illusions, Kahneman, 2012). Presence can be considered a continuously updated output from associative System-1 processing, whereas media awareness can be considered a continuously refreshed output from parallel propositional System-2 processing. This notion implies that both presence and media awareness can be quickly established. System-2 processing resulting in media awareness would not be more cognitively taxing or slower than System-1 processing resulting in presence (Gawronski and Bodenhausen, 2014). Next to the fact that both processes would be "default"-processes that are quickly established at the onset of media exposure, the notion of parallel-competitive processing presence would also suggest that both are largely independent processes that are not causally affecting each other. What evidence speaks for this assumption?

First, the idea of two causally unrelated processes would imply that media awareness does not affect presence. This idea converges well with the notion that perception (e.g., of optical illusions) is cognitively impenetrable (e.g., Sloman, 1996; Zeimbekis, 2015)—the perceptual impression is not affected by "better knowledge". If adapted to the present case, this principle would suggest that media awareness as a System 2-processing output does not directly alter perceptual presence sensations as a System-1 output. A user might not feel less present, simply because s/he gets more aware that "this is not really happening".[5] Likewise, being engaged in propositional System-2 processing should not interfere with being engaged in parallel associative System-1 processing.[6] Empirical evidence for this assumption is, however, scarce, indirect, and mixed. Two studies only indirectly illuminated if media awareness affects presence. Both studies did not directly measure media awareness, but manipulated consistency which we consider a trigger of media awareness. A recent experiment (Hofer et al., 2020) manipulated the semantic consistency of a VR environment (i.e., the external plausibility of an apartment) and found that these variations of plausibility did not affect users' sensation of spatial presence. Another experiment by Skarbez et al. (2018), Study 2) manipulated coherence based on the degree to which events in the VR environment adhered to laws of physics. This study, too, yielded no effects on different presence measures. However, in another recent study (Quaglia and Holecek, 2018), participants were subjected to a fear-of-height experience in VR. The authors found that virtual lucidity (i.e., "awareness that one is having a virtual experience", p. 1) was not only associated with lower fear and more daring behaviour in the presented "virtual plank"-scenario, but also lower spatial presence. In summary, in light of these scarce and mixed findings, the idea that media awareness does not affect presence remains a plausible

[5]We assume media awareness can only alter presence indirectly (e.g., if users, perhaps after a peak in media awareness, shift their attentional focus onto the medium interface or the real world and thus change the perceptual input that establishes the sensation of presence). But in the absence of these shifts in attention, media awareness might not alter the sensation of presence.

[6]This view can be further refined by recalling what Gawronski and Bodenhausen (2014) call the *operating principles* of System 1 and 2. The operating principle of System-1 or associative processing is that it works independently of truth judgments, whereas System-2 or propositional processing serves as the "validation of momentarily activated information on the basis of logical consistency" (p. 189). Adapted to the present case, in line with the operating principle of System-1, we believe presence occurs independent of users' media awareness (as a truth judgment), just like optical illusions usually occur despite better knowledge. However, in line with the operating principle of System-2, media awareness might invalidate sensations of presence, not by diminishing the perceptual sensation, but by invalidating perceptual beliefs emerging from this sensation such as that "this is really happening right now in front of me." (Sloman, 1996; Kahneman, 2012)—and subsequently by moderating effects of presence on the overall user experience. Our view thus converges with the operating principles of System 1 and 2 suggested by Gawronski and Bodenhausen (2014).

assumption which, however, needs to be further empirically scrutinized.

Second, the idea of two causally unrelated processes also implies that feeling present does not affect media awareness. Users' ability to engage in propositional System-2 processing and stay "media aware" might be unrelated to the extent that they engage in parallel associative System-1 processing or the intensity of their presence sensations. Admittedly, however, to date we know of no theoretical account or empirical study that would explicitly inform about this assumption.

Default-Interventionist Dual Processing

According to the default-interventionist dual processing logic (Evans and Stanovich, 2013), presence would be the outcome of quickly established, default, and continuously activated System-1 processing, while media awareness would be the outcome of slow, cognitively taxing and thus only occasionally activated System-2 processing. Following the default-interventionist logic, System-2 processes allow to intervene in System-1 processing and regulate (e.g., dismiss, weaken, contrast) related outcomes. Another typical characteristic of System-2 processing in a default-interventionist logic is that it is cognitively taxing. It requires working memory and attentional focus. Therefore, only if cognitive resources and motivation allow, perceptually-driven System-1 outputs might be overridden or regulated by effortful System-2 operations.

According to a default-interventionist logic, associative System-1 processing would be the default mode. Relatedly, feeling present (and temporarily "believing" in this sensation) would be a quick and default System-1 output in media exposure. However, users might engage in effortful System-2 processing to trigger media awareness and recall that "this is not really happening". According to the default-interventionist logic these interventions would require energy and would only be triggered if necessary. Only if sufficiently motivated and having sufficient cognitive capacity users might effortfully become media aware by accessing their higher-order propositional knowledge. Accordingly, following a default-interventionist logic, System-2 and System-1 processing, respectively media awareness and presence, would only occasionally co-occur in media exposure, namely when media awareness is effortfully triggered to causally affect presence.

What evidence speaks for these assumptions? We reviewed the mixed empirical evidence regarding a potential intervening influence of media awareness on presence above. Apart from these studies we are unaware of any direct empirical examinations of how media awareness and presence develop throughout exposure and potentially interact. Hence, we can only discuss the default-interventionist logic on theoretical grounds. A default-interventionist view on presence and media awareness has been implied by several authors in the literature, including ourselves in the past (e.g., Schubert, 2009; Hofer, 2016; Hartmann, 2017). In general, many media scholars argue that users approach media as "believers," and that perceiving the represented content rather than the representation is the default mode in media exposure. In contrast, users might engage in effortful evaluation of the representation (and hence become "media aware") only if this default mode encounters problems like inconsistencies, or because it triggers undesired psychological states like obnoxious fear (Gilbert et al., 1993; Grodal, 2002; Busselle and Bilandzic, 2008; Kahneman, 2012; Shapiro and Kim, 2012).

While these arguments speak for a default-interventionist notion, we would also like to highlight two potential problems with this notion, which make us skeptical that it provides a more suitable view on how media awareness and presence are related than the alternative parallel-competitive notion. First, it is unclear if recognizing a situation as a media exposure situation (i.e. both starting to be and staying media aware) is actually cognitively taxing, as the default-interventionist view would imply. At least, if compared to the typical cognitively taxing System-2 activity addressed in the literature, like for example solving mathematical problems or deeper analytical thinking (Tversky and Kahneman, 1974), activating and refreshing the belief that "this is not really happening" seems a relatively quick and effortless activity. However, we also call for future research that tests if this assumption is eventually correct and if, as we expect, staying "media aware" is not cognitively taxing, and also does not become more cognitively taxing with more intense presence sensations.

A second potential problem we see when applying the default-interventionist logic is that it would consider presence a quick default response, and media awareness a slow occasional interventionist activity. We admit that this view converges well with notions in the literature that believing is the default mode in media exposure, and disbelieving might take effort (Gilbert et al., 1993; Shapiro and Kim, 2012). We also agree that perceptual presence sensations are quickly established. Finally, we also admit that our view converges with a default-interventionist logic in assuming that the (System-2) belief "this is not really happening" can turn *more salient* if perceptual processing (default System-1) encounters inconsistencies (see **Figure 1** and **Figure 2**). However, in our view, presence is also *embedded* into basic media awareness (i.e., a knowing state "that this is not really happening"). In other words, contrary to the default-interventionist logic that argues presence precedes media awareness, we argue that basic media awareness precedes perceptual sensations like presence, because it is already triggered if a user consciously exposes him/herself to media technology. Similarly, the default-interventionist logic would suggest that feeling present (or "believing") would be the default mode in exposure, and media awareness the occasional intervention. Our approach, in contrast, suggests that both, perceptual sensations and media awareness are constantly co-occurring, and thus *jointly* define the *default* mode in media exposure.

To conclude, based on these arguments we think of presence and media awareness as two phenomena that are continuously co-occurring in VR exposure. This view converges with the notion of presence and media awareness as representing two parallel (competitive) processes. However, empirical research is needed to derive a more conclusive picture about how both processes co-occur and affect each other. Is it indeed not cognitively taxing, as we expect, to recall that "this is not really happening"? Is it indeed not harder, as we expect, for a

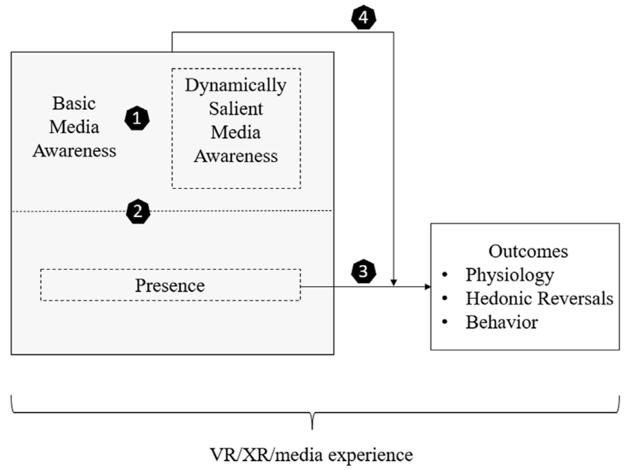

FIGURE 3 | Possible Interconnections between Presence and Media Awareness and Joint (Interactive) Effects on Outcomes that are Characterizing the Overall User Experience. 1) Media awareness consists of a basic and a dynamically varying part. Basic media awareness is triggered by (conscious) media exposure. Dynamically salient media awareness is triggered by factors outlined in **Figure 1**. 2) Media awareness and presence represent two parallel processes during exposure. We assume they might not causally affect each other (yet in the absence of empirical evidence this assumption remains speculative). 3) Presence affects outcomes, i.e., physiological (e.g., arousal), affective (e.g., fear), and behavioral (e.g., approach vs. withdrawal) aspects of the user experience. 4) We assume that basic and dynamically salient media awareness moderate these effects of presence on outcomes. For example, fear might be reversed into pleasurable excitation if media awareness reaches a certain salience level. Hence, users' overall media exposure experience needs to be explained based on the interaction of presence and media awareness.

user to stay media aware if presence intensifies? Does recalling that "this is not really happening" indeed not affect presence, as we expect? These central questions can only be more firmly answered based on future empirical evidence. While understanding the exact relations of media awareness and presence as two parallel processes requires further empirical scrutiny, we think the way both jointly shape the overall user experience can already be derived more firmly based on existing empirical research.

Contextualization: Media awareness Qualifies the *Consequences* of Presence

A core assumption of our approach is that media awareness provides the cognitive backdrop, or context, based on which immediate perceptual sensations like presence are subjectively interpreted by a user. Media awareness thus *qualifies*, *moderates*, or *contextualizes* effects of perceptual presence sensations on subsequent affective, cognitive, and behavioral responses in media exposure (see **Figure 3**).

Imagine a person sitting in a virtual living room in a perfect VR. This person might have exactly the same presence sensation as if she would be sitting in her real living room. Nevertheless, she might respond to the environment quite differently. The way her sensation of presence motivates subsequent psychological responses (e.g., arousal, emotions, thoughts, behaviour) might be strongly qualified by whether she believes the environment is mediated (thus attributing her presence sensation to technology) or real. Feeling present while believing "this is not really happening" is not the same, and does not cause the same consequences, as feeling present while believing "this is really happening". The person in the present example might, for instance, perceive that she is attacked by a bear in her living room. It will make a difference if the person, in parallel to this perceptual sensation, believes that this is really happening or not. Accordingly, media awareness, or "believing this is not really happening" matters, as this knowing state contextualizes perceptual sensations and changes their meaning (Berthiaume et al., 2021, p. 393), thus making it possible for users to respond to them differently.

More specifically, we propose that believing that "this is not really happening" tempers the effect of perceptual sensations as they seem *less self-relevant* (Abraham and von Cramon, 2009) and more *inconsequential* (i.e. unable to seriously physically or psychologically affect a person, Hartmann and Fox 2021). An analogy is to think of media awareness as a protective layer similar to a glass wall (e.g., when encountering a poisonous spider in a zoo, Russell, 1994; Gendler, 2019). The glass wall does not so much change the perceptual sensation of the object on the other side as it changes the overall meaning of the situation. By being aware of the glass wall when encountering the spider, the situation becomes less threatening, arousing, and perhaps even more exciting. This view on media awareness converges with the popular idea among scholars that media provide a protective layer (Andrade and Cohen, 2007), or playground (Vorderer, 2001), because represented objects have no physical impact (e.g., they do not hurt), and their psychological impact can be relatively well controlled (e.g., regulation of undesired affect).

We think the strongest empirical evidence for this view on media awareness comes from experimental studies showing that participants treat *identical* sensory stimuli, including VR, differently simply based on how they cognitively categorize the stimulus (e.g., avatar vs. agent, e.g., Ahn et al., 2012, mediated vs. real, Pönkänen et al., 2011). Other evidence comes from studies comparing real-life vs. virtual stimuli (e.g., Blankendaal et al., 2015; Gallup et al., 2019), although these studies potentially confound the manipulation of perceptual sensations (e.g., perceiving the stimulus as a physically embodied human-being) and people's higher-order cognitive belief or expectation (e.g., categorizing the other as an actually present human-being or representation).

Contextualization does not imply that we think presence would have no effect on users' overall experience, including emotions and behaviour. It implies that we think media awareness qualifies these effects. In fact, in line with a large body of evidence, we assume that in general, the stronger perceptual sensations like presence, the stronger their psychological consequences. For example, users in a fear-of-height VR that feel more spatially present should experience greater fear than users feeling less present in the environment. However, we believe that media awareness provides a decoupling from perceptual sensations like presence. The effect of perceptual sensations like presence on psychological consequences like fear might be less strong, and hence consequences might be weaker, the greater media awareness. For example, among users feeling equally present in a fear-of-height VR, those that are more aware "that this is not really happening" might feel less scared. Next to dampening the consequences of perceptual sensations, greater media awareness might also create novel opportunities on how to respond to perceptual sensations like "feeling present". For example, the stress of perceiving the physical presence of an aggressive bear might be reversed into excitement by being aware that the bear attack is not really happening.

More specifically, we suggest that both basic and dynamically salient media awareness qualify the overall user experience in the following ways (see **Figure 3**, see also Hartmann and Fox, 2021):

- *Media awareness reduces subsequent physiological responses, arousal and affect.* Perceptual sensations like presence might precede and inform physiological and affective responses (i.e., people affectively respond to what they perceive). However, media awareness might weaken the coupling and impact of perceptual sensations on these responses. In general, recalling that "this is not really happening" is considered an effective way to regulate undesired affect in media exposure (Cantor and Wilson, 1988; Hofer et al., 2015). However, media awareness is not just a coping strategy. In an experiment by Pönkänen et al. (2011), participants were exposed to an identical stimulus of another person's animated face. One group was made to believe that the face was a picture on a screen, the other group believed it to be the head of somebody looking through a window from the adjacent room. Eye-gazing of the other person triggered less arousal among the group that believed the face was just a picture, as compared to the other group that believed seeing a real person (see Risko et al., 2016, for related findings). Relatedly, van der Waal et al. (2021) examined how people respond to food stimuli (e.g., chocolate) in real life vs. VR. They find that exposure to food vs. non-food stimuli leads to more salivation in participants in real life, but not in VR. Potentially, awareness "that this is not real" suppressed users' salvation response. Another study by Quaglia and Holecek (2018) found that participants in a fear-of-height VR reported less fear, the more they stayed aware that they were immersed in a VR application. In summary, these studies suggest that (greater) media awareness might weaken the effect of presence on physiological responses (e.g., salivation, arousal) and emotions (e.g., fear).
- *Media awareness triggers hedonic reversals.* Media awareness is known to also allow for hedonic reversals in which negative primary affect is reappraised as something positive. Hence, the consequences of perceptual sensations (e.g., sensing the presence of an attacking bear) might not only be dampened but also reversed in their valence. For example, while sensing the presence of an attacking bear should instigate distress, being aware that "this is not really happening" allows to reverse the valence of this arousal, and thus turn distress into pleasurable excitement (Andrade and Cohen, 2007). This principle of hedonic reversals is well known from roller-coaster rides (body in fear, mind believes it is safe = fun) and other pleasurable body-over-mind experiences (e.g., chili consumption, Rozin et al., 2013). Hedonic reversals seem to work even if encountering highly immersive representations. For example, in the fear-of-height VR study by Quaglia and Holecek (2018) participants indeed enjoyed the fearful sensation more, the more they stayed media aware.
- *Media awareness instigates more daring, exploratory, and playful behavior.* If media awareness makes users recognize the situation as less consequential, and thus less threatening or risky than it seems, it is plausible that they adapt their behavior accordingly. Users might be inclined to engage in more exploratory, risky, and daring behaviour than they would if they were less media aware, or if they believed to be present in a real-world situation. For example, in the fear-of-height VR study by Quaglia and Holecek (2018), in which participants had to step on a plank, participants that were more media aware were more likely to dare jumping off the plank. This finding converges well with the idea of media providing a safe playground in which users explore boundaries (Vorderer, 2001), for better (e.g., entertainment, training) or worse (e.g., disinhibited harmful behaviour, like harassment or trolling).

CONCLUSION AND FUTURE RESEARCH AGENDA: HOW DO USERS EXPERIENCE VR?

In the present paper we aimed to conceptualize the typical VR experience. This goal was inspired by Slater's (2009) widely recognized approach, and particularly his notion of a plausibility illusion, and Skarbez et al. (2017) revision of the plausibility illusion as coherence. We embedded the concept of plausibility into a larger, more general model on how individuals might experience and respond to (mediated) representations, including highly immersive VR content. In our approach, both media awareness and presence are key concepts. We drew on interdisciplinary literature (particularly on the dualistic nature of representations) to draft a parallel-process account, in which the perceptual sensation of presence is contextualized by users' media awareness. We argued (see **Figure 3**) that both processes jointly shape the overall user experience, because media awareness moderates the effects of presence on physiological responses, affect, and behaviour. In our view, media awareness consists of basic media awareness that is initiated when a user consciously starts a media exposure episode. Therefore, we argue that, to a certain extent, users are constantly media aware. In addition, however, media awareness can vary in salience above these baseline levels throughout exposure. Integrating previous ideas about a plausibility illusion (Slater, 2009) and coherence (Skarbez et al., 2015), we argued that perceived (im)plausibility, both on a sensory/affordance and semantic level, next to motivated recall, affects this dynamic part of media awareness. We reckoned that familiarity might reduce this dynamic media awareness. We also engaged in the thorny topic of how perceptual presence sensations and media awareness might be mutually related, and propose that both co-occur in parallel during exposure without causally affecting each other. Stressing the importance of media awareness, we concluded with a view on how both interact to jointly affect physiological, affective, and behavioral responses that characterize the overall user experience. Altogether, the provided conceptualization suggests that media awareness, albeit often neglected in theoretical explications—and perhaps particularly in literature on VR that often emphasizes users' life-like responses—matters. Based on the present approach we believe that the user experience of VR, and arguably of any presence-evoking medium, can only be thoroughly understood if both presence and media awareness are jointly taken into account.

The proposed parallel-processing account answers related calls in the literature and promises to clarify important prevailing questions. For example, Pan and Hamilton (2018), p. 3) recently argued in a special issue on VR as a tool to study social behaviour that "little is known about the cognitive processes which allow us to engage in dual realities." Indeed, in most articles about VR or related media, scholars mention users' media awareness or users' knowledge "that this is not really happening" only as a curious side aspect. Other related approaches suggested that users stay mindless during media exposure, and are hence deluded (e.g., "media equation," Reeves and Nass, 1996). With the present account, however, we hope we challenged these views to potentially make room for a more fine-grained understanding of the dual realities of media exposure in general, and users' VR experience in particular.

A skeptical reader might wonder how our emphasis on media awareness aligns with the evidence that users appear to respond to VR as if it was real? Our account of media awareness is of course not neglecting that particularly immersive media like VR might trigger perceptual sensations coupled with psychological responses that look very similar to responses we observe to equivalent real-world stimuli. We think that these responses happen, because, as we discussed, presence is effective, and changes in the environment can effectively trigger changes in users, if users feel present. For instance, in VR, just like in real life, standing on a small plank very high above the ground evokes more fear as compared to standing on the ground, because users feel present. Hartmann (2017) suggested addressing these "lifelike responses" as *structurally equivalent* responses (the structure of responses to mediated condition A vs. B is equivalent to the responses to A and B we would observe in real life). However, as seen in examples of hedonic reversals (users are more likely to enjoy standing on a high skyscraper in VR) or very risky behaviour (users dare to jump off the skyscraper in VR), even in VR structural equivalence is not a given. Furthermore, in almost all studies we observe significant differences in the *intensity* of responses. The difference in response to condition A vs. B, as well as the overall intensity level, is in almost any study on mediated exposure, including VR, strikingly lower than responses to equivalent real-life conditions (e.g., Blankendaal et al., 2015). Our account suggests that this is largely a result of users' media awareness that is either dampening or even reversing these responses to conditions A vs. B.

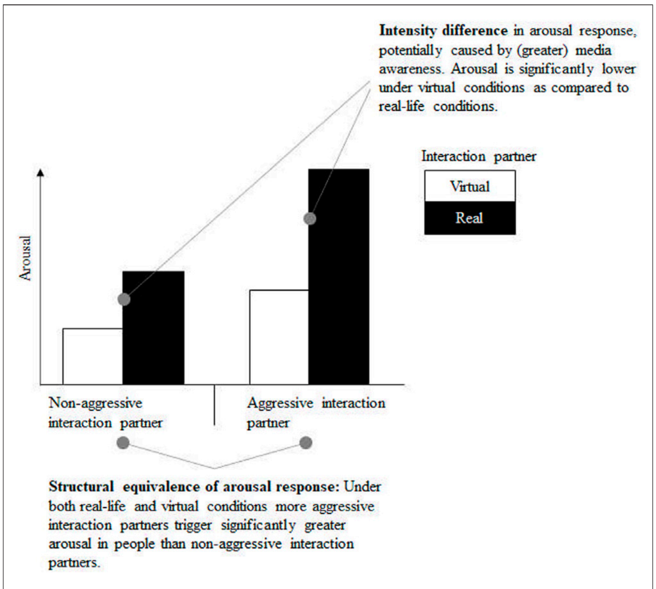

FIGURE 4 | Structural equivalence vs. intensity difference of user responses—illustrated in the context of findings by Blankendaal et al. (2015).

To summarize, **Figure 4** illustrates the idea of structural equivalence vs. intensity difference of people's responses under real vs. virtual conditions in the context of the findings of Blankendaal et al. (2015). Their study found that people are more aroused if the interaction partner behaves aggressively vs. peacefully, and that this is equally true for real and virtual interaction partners. Yet, the study also found clear differences in the intensity of people's arousal (both under peaceful and aggressive conditions), with people interacting with a virtual agent being remarkably less aroused than people interacting with a real confederate in the lab. We would argue that people were more aroused if confronted with an aggressive vs. peaceful real or virtual interaction partner, because they intuitively perceived the interaction partner to be co-present. At the same time, we assume that users' greater awareness that the virtual partner was actually not really co-present dampened arousal under virtual conditions.

Similar to Gonzalez-Franco and Lanier (2017), e.g., p. 7), our attempt to conceptualize the overall VR experience of course also raises many open questions. But we think our explications directly suggest testable propositions or hypotheses.

First, to more fully understand the VR user experience in the future, research examining how media awareness, including plausibility cues, and presence sensations relate to each other seems important. For example, future studies could test our model's assumptions by examining the proposed cognitive impenetrability of presence. Scholars could also test if keeping media awareness salient is affected by perceptual sensations like presence, and whether or not it is a cognitively taxing activity (perhaps depending on the intensity of presence).

Relatedly, mostly for the sake of simplicity, but also because these subtypes might be correlated, we treated presence as a unitary concept in this article and did not discuss commonly distinguished subtypes like spatial, social, and self-presence (Lee, 2004). However, future studies could discuss and test if our assumptions generally apply to all types of presence. For example, is recalling that "this is not truly happening" indeed similarly effortless when feeling spatially or socially present or self-present? Or, to raise another question, are spatial, social, and self-presence indeed equally cognitively impenetrable and thus unaffected by parallel media awareness?

Second, and not less importantly, researchers might want to test our propositions regarding the joint effects of presence and media awareness on arousal, affect, or behaviour. For example, it would be intriguing to address this hypothesis by manipulating media awareness under conditions of high social presence. Adapting a design of Pönkänen et al. (2011), participants could be exposed to a scene played by real human confederates behind a window. The experimental manipulation could consist of convincing one experimental group that what they see is a sophisticated VR simulation. The other group sees the exact same scene, but is convinced that they simply observe people in the adjacent room. Measures of presence, physiological responses, subjective or behavioral measures could serve as dependent variables. In addition, a yet to be developed measure of media awareness would have to be included as well. Such a design would allow to test if variations in media awareness, even if encountering "perfect stimuli," qualify users' responses.

Third, we think future research should test the cues that we propose to trigger the belief that "this is not really happening." Further insight into when people believe that "this is not real," also in non-mediated situations (Timmins and Lombard, 2005), might help positioning VR more clearly as "another environmental stimulus" (like a picture or a real object), that triggers general psychological mechanisms linked to reality perception and sense-making in a specific way, resulting in a specific user experience. In this context, our approach should also be merged with other approaches that generally place (in)consistencies and plausibility, respectively the extent perceived objects meet users' expectations, at the heart of presence (Seth et al., 2012) and related user experiences (Latoschik and Wienrich, 2021). While we focused on (in)consistencies that potentially trigger media awareness in this paper, certain (in)consistencies such as profound violations of basic laws of spatial perception might also affect presence experiences (Hofer et al., 2020).

Fourth, as a methodological challenge, future research would require a separate assessment of the basic vs. dynamic part of media awareness. A related psychometrically tested and validated measurement does not exist to date. A valid, reliable and time-sensitive measurement of media awareness might also require a solid theoretical understanding of its dynamic variation throughout exposure. In the light of recent conceptualizations of the temporal development of perceptual processes (e.g., Wirth et al., 2007; Merfeld et al., 2016), does media awareness shift in a dichotomous way (from activated to non-activated) or in a more continuous way (Merfeld et al., 2016, see similarly; Skarbez et al., 2018)?

Fifth and finally, we think the present approach might also raise questions about how users experience mediated representations in general (e.g., see Wolf, 2017). For instance, are the same mechanisms at play when we perceive robots as socially present beings? The above-mentioned findings by Pönkänen et al. (2011) might suggest this. Relatedly, we wonder if the same processes like the one we sketch in the present approach are at work in exposure to symbolic (text) vs. analog (e.g., audio-visual) stimuli. Does media awareness, for example, also qualify users' response to a text like "you are attacked by a bear!", as we assume it does qualify the response in VR exposure? If these processes converge, it would be a bold yet certainly deserving endeavour to work towards a general model of how users experience media or mediated representations.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

TH and MH have collaborated on this manuscript. The order of the authors reflects the amount of work put into the manuscript. Specifically, TH: Original idea of the concept, most of the writing; MH: Contributing and editing.

ACKNOWLEDGMENTS

We would like to thank both reviewers as well as Jesse Fox and Jozua Murris for their insightful feedback on earlier versions of this manuscript.

REFERENCES

Abraham, A., and von Cramon, D. Y. (2009). Reality = Relevance? Insights from Spontaneous Modulations of the Brain's Default Network when Telling Apart Reality from Fiction. *PLoS One* 4 (3), e4741. doi:10.1371/journal.pone.0004741

Ahn, S. J., Fox, J., and Bailenson, J. N. (2012). "Avatars," in *Leadership in Science and technology.A Reference Handbook*. Editor W. S. Bainbridge (Thousand Oaks, CA: SAGE Publications, Inc), 695–702.

Andrade, E. B., and Cohen, J. B. (2007). On the Consumption of Negative Feelings. *J. Consum Res.* 34, 283–300. doi:10.1086/519498

Berthiaume, M., Corno, G., Nolet, K., and Bouchard, S. (2021). A Novel Integrated Information Processing Model of Presence. *PRESENCE: Virtual and Augmented Reality* 27, 378–399. doi:10.1162/pres_a_00336

Biocca, F. (1997). The Cyborg's Dilemma: Progressive Embodiment in Virtual Environments. *J. Comput.-Mediat. Commun.* 3 (2). doi:10.1111/j.1083-6101.1997.tb00070.x

Biocca, F., Kim, J., and Choi, Y. (2001). Visual Touch in Virtual Environments: An Exploratory Study of Presence, Multimodal Interfaces, and Cross-Modal Sensory Illusions. *Presence: Teleoperators & Virtual Environments* 10 (3), 247–265. doi:10.1162/105474601300343595

Blankendaal, R., Bosse, T., Gerritsen, C., de Jong, T., and de Man, J. (2015). "Are Aggressive Agents as Scary as Aggressive Humans?," in Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems, 553–561.

Busselle, R., and Bilandzic, H. (2008). Fictionality and Perceived Realism in Experiencing Stories: A Model of Narrative Comprehension and Engagement. *Commun. Theor.* 18, 255–280. doi:10.1111/j.1468-2885.2008.00322.x

Camos, V., Johnson, M., Loaiza, V., Portrat, S., Souza, A., and Vergauwe, E. (2018). What is Attentional Refreshing in Working Memory?. *Ann. N Y Acad. Sci.* 1424 (1), 19–32. doi:10.1111/nyas.13616

Cantor, J., and Wilson, B. J. (1988). Helping Children Cope with Frightening media Presentations. *Curr. Psychol.* 7 (1), 58–75. doi:10.1007/BF02686664

Cowan, N. (2017). The many Faces of Working Memory and Short-Term Storage. *Psychon. Bull. Rev.* 24 (4), 1158–1170. doi:10.3758/s13423-016-1191-6

Cupchik, G. C. (2001). Theoretical Integration Essay: Aesthetics and Emotion in Entertainment Media. *Media Psychol.* 3 (1), 69–89. doi:10.1207/s1532785xmep0301_04

de Gelder, B., Kätsyri, J., and de Borst, A. W. (2018). Virtual Reality and the New Psychophysics. *Br. J. Psychol.* 109 (3), 421–426. doi:10.1111/bjop.12308

DeLoache, J. S., Chiong, C., Sherman, K., Islam, N., Vanderborght, M., Troseth, G. L., et al. (2010). Do babies Learn from Baby media? *Psychol. Sci.* 21 (11), 1570–1574. doi:10.1177/0956797610384145

DeLoache, J. S., Pierroutsakos, S. L., and Uttal, D. H. (2003). The Origins of Pictorial Competence. *Curr. Dir. Psychol. Sci.* 12 (4), 114–118. doi:10.1111/1467-8721.01244

Dokic, J., and Martin, J.-R. (2017). Felt Reality and the Opacity of Perception. *Topoi* 36 (2), 299–309. doi:10.1007/s11245-015-9327-2

Evans, J. S. B. T. (2007). On the Resolution of Conflict in Dual Process Theories of Reasoning. *Thinking & Reasoning* 13 (4), 321–339. doi:10.1080/13546780601008825

Evans, J. S. B. T., and Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition. *Perspect. Psychol. Sci.* 8 (3), 223–241. doi:10.1177/1745691612460685

Flavell, J. H. (2000). Development of Children's Knowledge about the Mental World. *Int. J. Behav. Dev.* 24 (1), 15–23. doi:10.1080/016502500383421

Flavell, J. H., Flavell, E. R., and Green, F. L. (1983). Development of the Appearance-Reality Distinction. *Cogn. Psychol.* 15 (1), 95–120. doi:10.1016/0010-0285(83)90005-1

Fox, J., Arena, D., and Bailenson, J. N. (2009). Virtual Reality A Survival Guide For The Social Scientist. *J. Media Psychol.* 21 (3), 95–113. doi:10.1027/1864-1105.21.3.95

Frey, F. (2018). The Experiential Mode of media Reception: A Holistic Framework Concept. *Commun. Theor.* 28 (4), 487–510. doi:10.1093/ct/qty010

Gallup, A. C., Vasilyev, D., Anderson, N., and Kingstone, A. (2019). Contagious Yawning in Virtual Reality Is Affected by Actual, but Not Simulated, Social Presence. *Sci. Rep.* 9 (1), 294. doi:10.1038/s41598-018-36570-2

Gawronski, B., and Bodenhausen, G. V. (2011). The Associative-Propositional Evaluation Model. *Adv. Exp. Soc. Psychol.* 44, 59–127. doi:10.1016/B978-0-12-385522-0.00002-0

Gawronski, B., and Bodenhausen, G. V. (2014). "The Associative-Propositional Evaluation Model: Operating Principles and Operating Conditions of Evaluation," in *Dual-process Theories of the Social Mind* (New York, NY: The Guilford Press), 188–203.

Gendler, T. (2019). "Alief and Belief," in *Contemporary Epistemology*. Editors J. Fantl, M. McGrath, and E. Sosa (Hoboken, NJ: John Wiley & Sons), 91–109. doi:10.1002/9781119420828.ch7

Gendler, T. S. (2008). Alief and Belief. *J. Philos.* 105 (10), 634–663. doi:10.5840/jphil20081051025

Gilbert, D. T., Tafarodi, R. W., and Malone, P. S. (1993). You Can't Not Believe Everything You Read. *J. Personal. Soc. Psychol.* 65 (2), 221–233. doi:10.1037/0022-3514.65.2.221

Gilbert, S. B. (2016). Perceived Realism of Virtual Environments Depends on Authenticity. *Presence: Teleoperators and Virtual Environments* 25 (4), 322–324. doi:10.1162/PRES_a_00276

Gonzalez-Franco, M., and Lanier, J. (2017). Model of Illusions and Virtual Reality. *Front. Psychol.* 8, 1125. doi:10.3389/fpsyg.2017.01125

Grodal, T. (2002). "The Experience of Realism in Audiovisual Representation," in *Realism and 'reality' in Film and media*. Editor A. Jerslev (Copenhagen: Museum Tusculanum Press, University of Copenhagen), 67–91.

Grodal, T. (2006). The PECMA Flow: A General Model of Visual Aesthetics. *Film Stud.* 8 (1), 1–11. doi:10.7227/FS.8.3

Haans, A., and IJsselsteijn, W. A. (2012). Embodiment and Telepresence: Toward a Comprehensive Theoretical Framework. *Interacting Comput.* 24 (4), 211–218. doi:10.1016/j.intcom.2012.04.010

Hall, A. (2003). Reading Realism: Audiences' Evaluations of the Reality of Media Texts. *J. Commun.* 53, 624–641. doi:10.1111/j.1460-2466.2003.tb02914.x

Hartmann, T., and Fox, J. (2021). "Entertainment in Virtual Reality and beyond," in *The Oxford Handbook of Entertainment Theory*. Editors P. Vorderer and C. Klimmt (Oxford, England: Oxford University Press), 716–732. doi:10.1093/oxfordhb/9780190072216.013.37

Hartmann, T. (2012). "Players' Experiential and Rational Processing of Virtual Violence," in *Vice City virtue.Moral Issues in Digital Game Play*. Editors K. Poels and S. Malliet. First edition (Leuven: Acco), 135–150.

Hartmann, T. (2017). The "Moral Disengagement in Violent Videogames" Model. *Game Stud.* 17 (2). Available at: http://gamestudies.org/1702/articles/hartmann.

Hartmann, T., Wirth, W., Vorderer, P., Klimmt, C., Schramm, H., and Böcking, S. (2015). "Spatial Presence Theory: State of the Art and Challenges Ahead," in *Immersed in media.Telepresence Theory, Measurement & Technology*. Editors M. Lombard, F. Biocca, J. Freeman, W. A. Ijsselsteijn, and R. J. Schaevitz (New York, NY: Routledge), 115–135. doi:10.1007/978-3-319-10190-3_7

Herschbach, M. (2015). Direct Social Perception and Dual Process Theories of Mindreading. *Conscious. Cogn.* 36, 483–497. doi:10.1016/j.concog.2015.04.001

Hofer, M., Burkhard, L., and Allemand, M. (2015). Age Differences in Emotion Regulation during a Distressing Film Scene. *J. Media Psychol.* 27 (1), 47–52. doi:10.1027/1864-1105/a000134

Hofer, M., Hartmann, T., Eden, A., Ratan, R., and Hahn, L. (2020). The Role of Plausibility in the Experience of Spatial Presence in Virtual Environments. *Front. Virtual Real.* 1, 1–9. doi:10.3389/frvir.2020.00002

Hofer, M. (2016). *Involvement und Presence*. Baden-Baden: Nomos.

ISPR (2001). What Is Presence? Available at: https://ispr.info/about-presence-2/about-presence/ (Accessed March 2, 2018).

Jacobs, C., and Silvanto, J. (2015). How Is Working Memory Content Consciously Experienced? the 'conscious Copy' Model of WM Introspection. *Neurosci. Biobehavioral Rev.* 55, 510–519. doi:10.1016/j.neubiorev.2015.06.003

Kahneman, D., and Frederick, S. (2002). "Representativeness Revisited: Attribute Substitution in Intuitive Judgment," in *Heuristics and Biases*. Editors T. Gilovich, D. Griffin, and D. Kahneman. 1st ed. (Cambridge, England: Cambridge University Press), 49–81. doi:10.1017/CBO9780511808098.004

Kahneman, D. (2012). *Thinking, Fast and Slow*. London: Penguin Books.

Krcmar, M., and Eden, A. (2019). Rational versus Intuitive Processing. *J. Media Psychol.* 31 (1), 2–11. doi:10.1027/1864-1105/a000215

Latoschik, M. E., and Wienrich, C. (2021). Coherence and Plausibility, Not Presence?! Pivotal Conditions for XR Experiences and Effects, a Novel Model. *arXiv*. Available at: http://arxiv.org/abs/2104.04846 (Accessed September 21, 2021).

Lee, K. M. (2004). Presence, Explicated. *Commun. Theor.* 14, 27–50. doi:10.1111/j.1468-2885.2004.tb00302.x

Lombard, M., and Ditton, T. (1997). At the Heart of it All: The Concept of Presence. *J. Computer-Mediated Commun.* 3 (2). Available at: http://209.130.1.169/jcmc/vol3/issue2/lombard.html.

Merfeld, D. M., Clark, T. K., Lu, Y. M., and Karmali, F. (2016). Dynamics of Individual Perceptual Decisions. *J. Neurophysiol.* 115 (1), 39–59. doi:10.1152/jn.00225.2015

Michelle, C. (2007). Modes of Reception: A Consolidated Analytical Framework. *Commun. Rev.* 10 (3), 181–222. doi:10.1080/10714420701528057

Nanay, B. (2005). Is Twofoldness Necessary for Representational Seeing? *The Br. J. Aesthetics* 45 (3), 248–257. doi:10.1093/aesthj/ayi034

Nieding, G., Ohler, P., Diergarten, A. K., Möckel, T., Rey, G. D., and Schneider, W. (2017). The Development of Media Sign Literacy-A Longitudinal Study with 4-Year-Old Children. *Media Psychol.* 20 (3), 401–427. doi:10.1080/15213269.2016.1202773

Pan, X., and Hamilton, A. F. d. C. (2018). Why and How to Use Virtual Reality to Study Human Social Interaction: The Challenges of Exploring a New Research Landscape. *Br. J. Psychol.* 109 (3), 395–417. doi:10.1111/bjop.12290

Pönkänen, L. M., Alhoniemi, A., Leppänen, J. M., and Hietanen, J. K. (2011). Does it Make a Difference if I Have an Eye Contact with You or with Your Picture? an ERP Study. *Soc. Cogn. Affective Neurosci.* 6 (4), 486–494. doi:10.1093/scan/nsq068

Popova, L. (2010). *Perceived Reality of media Messages*. Santa Barbara: University of California.

Quaglia, J. T., and Holecek, A. (2018). "Lucid Virtual Dreaming: Antecedents and Consequents of Virtual Lucidity during Virtual Threat," in 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), 1–8. doi:10.1109/VR.2018.8446546

Rebenitsch, L., and Owen, C. (2016). Review on Cybersickness in Applications and Visual Displays. *Virtual Reality* 20 (2), 101–125. doi:10.1007/s10055-016-0285-9

Reeves, B., and Nass, C. I. (1996). *The media Equation*. Stanford, CA: CSLI Publications.

Risko, E. F., Richardson, D. C., and Kingstone, A. (2016). Breaking the Fourth Wall of Cognitive Science. *Curr. Dir. Psychol. Sci.* 25 (1), 70–74. doi:10.1177/0963721415617806

Ross, M. (2015). *3D Cinema: Optical Illusions and Tactile Experiences*. New York: Palgrave Macmillan.

Rozin, P., Guillot, L., Fincher, K., Rozin, A., and Tsukayama, E. (2013). Glad to Be Sad, and Other Examples of Benign Masochism. *Judgment Decis. Making* 8 (4), 439–447.

Russell, J. A. (1994). Is There Universal Recognition of Emotion from Facial Expression? A Review of the Cross-Cultural Studies. *Psychol. Bull.* 115 (1), 102–141. doi:10.1037/0033-2909.115.1.102

Schlottmann, A. (2001). Children's Probability Intuitions: Understanding the Expected Value of Complex Gambles. *Child. Dev.* 72 (1), 103–122. doi:10.1111/1467-8624.00268

Schubert, T. W. (2009). A New Conception of Spatial Presence: Once Again, with Feeling. *Commun. Theor.* 19 (2), 161–187. doi:10.1111/j.1468-2885.2009.01340.x

Senju, A., and Johnson, M. H. (2009). The Eye Contact Effect: Mechanisms and Development. *Trends Cogn. Sci.* 13 (3), 127–134. doi:10.1016/j.tics.2008.11.009

Seth, A. K., Suzuki, K., and Critchley, H. D. (2012). An Interoceptive Predictive Coding Model of Conscious Presence. *Front. Psychol.* 2, 395. doi:10.3389/fpsyg.2011.00395

Shapiro, M. A., and Kim, H. (2012). Realism Judgments and Mental Resources: A Cue Processing Model of Media Narrative Realism. *Media Psychol.* 15, 93–119. doi:10.1080/15213269.2011.649666

Skarbez, R., Brooks, F. P., and Whitton, M. C. (2018). "Immersion and Coherence in a Stressful Virtual Environment," in Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology, 1–11. doi:10.1145/3281505.3281530

Skarbez, R., Neyret, S., Brooks, F. P., Slater, M., and Whitton, M. C. (2017). A Psychophysical experiment Regarding Components of the Plausibility Illusion. *IEEE Trans. Vis. Comput. Graphics* 23 (4), 1369–1378. doi:10.1109/TVCG.2017.2657158

Skarbez, R. (2016). *Plausibility Illusion in Virtual Environments*. Chapel Hill, NC: The University of North Carolina at Chapel Hill University Libraries. doi:10.17615/2MG3-GH93

Slater, M. (2009). Place Illusion and Plausibility Can lead to Realistic Behaviour in Immersive Virtual Environments. *Phil. Trans. R. Soc. B* 364 (1535), 3549–3557. doi:10.1098/rstb.2009.0138

Slater, M., Antley, A., Davison, A., Swapp, D., Guger, C., Barker, C., et al. (2006). A Virtual Reprise of the Stanley Milgram Obedience Experiments. *PLoS One* 1 (1), e39. doi:10.1371/journal.pone.0000039

Slater, M., and Wilbur, S. (1997). A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. *Presence: Teleoperators & Virtual Environments* 6, 603–616. doi:10.1162/pres.1997.6.6.603

Sloman, S. A. (1996). The Empirical Case for Two Systems of Reasoning. *Psychol. Bull.* 119 (I), 3–22. doi:10.1037/0033-2909.119.1.3

Smith, E. R., and DeCoster, J. (20020). Dual-Process Models in Social and Cognitive Psychology: Conceptual Integration and Links to Underlying Memory Systems. *Pers. Soc. Psychol. Rev.* 40 (2), 108–131. doi:10.1207/S15327957PSPR0402_01

Suckfüll, M., and Scharkow, M. (2009). Modes of Reception for Fictional Films. *Communications* 34 (4), 361–384. doi:10.1515/COMM.2009.023

Timmins, L. R., and Lombard, M. (2005). When "Real" Seems Mediated: Inverse Presence. *Presence: Teleoperators & Virtual Environments* 14 (4), 492–500. doi:10.1162/105474605774785307

T. Koblížek (Editor) (2017). *The Aesthetic Illusion in Literature and the Arts* (London, England: Bloomsbury Academic).

Turner, P. (2016). Presence: Is it Just Pretending? *AI Soc.* 31 (2), 147–156. doi:10.1007/s00146-014-0579-y

Tversky, A., and Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science* 185 (4157), 1124–1131. doi:10.1126/science.185.4157.1124

van der Waal, N. E., Janssen, L., Antheunis, M., Culleton, E., and van der Laan, L. N. (2021). The Appeal of Virtual Chocolate: A Systematic Comparison of Psychological and Physiological Food Cue Responses to Virtual and Real Food. *Food Qual. Preference* 90, 104167. doi:10.1016/j.foodqual.2020.104167

Vorderer, P. (2001). It's All Entertainment-Sure. But what Exactly Is Entertainment? Communication Research, media Psychology, and the Explanation of Entertainment Experiences. *Poetics* 29, 247–261. doi:10.1016/S0304-422X(01)00037-7

Waterworth, J., and Tjostheim, I. (2021). Integrating the World of Presence Theory: Illusion, Pretence, Attending, and Pretending. *PsyArXiv*, 1–30. doi:10.31234/osf.io/qbcfe

Wirth, W., Hartmann, T., Böcking, S., Vorderer, P., Klimmt, C., Schramm, H., et al. (2007). A Process Model of the Formation of Spatial Presence Experiences. *Media Psychol.* 9, 493–525. doi:10.1080/15213260701283079

Wolf, W. (2017). "Aesthetic Illusion(s)?," in *The Aesthetic Illusion in Literature and the Arts*. Editor T. Koblížek (London, England: Bloomsbury Academic).

Wolf, W. (2014). "Illusion (Aesthetic)," in *The Living Handbook of Narratology*. Editors P. Hühn, J. Pier, W. Schmid, and J. Schönert. Available at: https://www.lhn.uni-hamburg.de/node/71.html. doi:10.1515/9783110316469.270

Wollheim, R. (1998). Symposium: Wollheim on Pictorial Representation. *The J. Aesthetics Art Criticism* 56 (3), 217–226. doi:10.2307/43236110.1111/1540_6245.jaac56.3.0217

Zeimbekis, J. (2015). "Seeing, Visualizing, and Believing," in *The Cognitive Penetrability of Perception*. Editors J. Zeimbekis and A. Raftopoulos (Oxford, England: Oxford University Press), 298–327. doi:10.1093/acprof:oso/9780198738916.003.0013

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

PERSPECTIVE
published: 16 June 2022
doi: 10.3389/frvir.2022.694433

Congruence and Plausibility, Not Presence: Pivotal Conditions for XR Experiences and Effects, a Novel Approach

Marc Erich Latoschik[1*†] *and Carolin Wienrich*[2†]

[1]*Human-Computer Interaction Group, University of Würzburg, Würzburg, Germany,* [2]*Human-Technology-Systems Group, University of Würzburg, Würzburg, Germany*

Presence is often considered the most important quale describing the subjective feeling of being in a computer-generated and/or computer-mediated virtual environment. The identification and separation of orthogonal presence components, i.e., the place illusion and the plausibility illusion, has been an accepted theoretical model describing Virtual Reality (VR) experiences for some time. This perspective article challenges this presence-oriented VR theory. First, we argue that a place illusion cannot be the major construct to describe the much wider scope of virtual, augmented, and mixed reality (VR, AR, MR: or XR for short). Second, we argue that there is no plausibility illusion but merely plausibility, and we derive the place illusion caused by the congruent and plausible generation of spatial cues and similarly for all the current model's so-defined illusions. Finally, we propose congruence and plausibility to become the central essential conditions in a novel theoretical model describing XR experiences and effects.

Keywords: XR, experience, presence, congruence, plausibility, coherence, theory, prediction

OPEN ACCESS

Edited by:
Doug A. Bowman,
Virginia Tech, United States

Reviewed by:
Anderson Maciel,
Federal University of Rio Grande do Sul, Brazil
Mark Billinghurst,
University of South Australia, Australia

****Correspondence:***
Marc Erich Latoschik
marc.latoschik@uni.wuerzburg.de

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Virtual Reality and Human Behavior, a section of the journal Frontiers in Virtual Reality

Received: *13 April 2021*
Accepted: *04 May 2022*
Published: *16 June 2022*

Citation:
Latoschik ME and Wienrich C (2022) Congruence and Plausibility, Not Presence: Pivotal Conditions for XR Experiences and Effects, a Novel Approach. Front. Virtual Real. 3:694433. doi: 10.3389/frvir.2022.694433

INTRODUCTION

"*A review and categorization of definitions of presence has demonstrated that it is an unusually rich and diverse concept. [. . .] Presence, and definitions of presence, touch on profound issues involving the nature of reality and existence; human cognition, affect, and perception; the characteristics, uses, and impacts of primitive, advanced, and futuristic technologies; and the subtleties of interpersonal communication and human–technology interaction*" (Lombard and Jones, 2015, 30).

Lombard and Jones highlight the significance of the presence construct. However, they also reflect on the wide scope, the potential diversity of definitions, and hence the blurred concreteness of its very nature. There are other considerable problems with the presence construct. Biocca's book problem addresses the technology-driven interpretation since presence can be experienced by imagination and/or in narratives presented in nonimmersive media such as books (Schubert and Crusius, 2002). Then, presence models often expose a sole dependency on other qualia and constructs such as the place, plausibility, and social presence illusions (Skarbez et al., 2017) or the virtual body ownership illusion (Latoschik et al., 2017; Waltemate et al., 2018). Even more, a central focus on a sense of "*being there*" for XR applications does not capture the essence of the many variations of XR covered by the Virtuality Continuum (Milgram and Kishino, 1994). In essence, if we want to guide designers and developers to create compelling XR applications and experiences as initially motivated by Heeter,

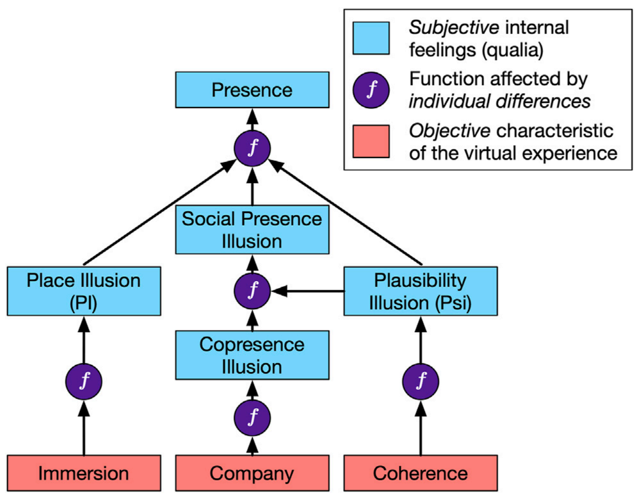

FIGURE 1 | Relationships between presence concepts as proposed by Skarbez et al. (2017, 96:23); layout redesigned by the authors and framed as "*Skarbez-model*".

(1992), we need well-defined qualities to strive for, with pragmatic ways to operationalize modifications to these qualities and provide clear-cut entry-points for a user-centered design process.

RELATED WORK

There now is a considerable body of knowledge on presence, see excellent overviews in the study by Lombard et al., (2015); Skarbez et al., (2017). We follow the study by Lombard and Jones, (2015) and start by defining **presence: The related quale mediated by XR-technology, that is, the degree one believes that she exists within a mediated space** (Jerome and Jordan, 2007), including concepts of virtual presence (Heeter, 1992) and telepresence: "*The biggest challenge to developing telepresence is achieving that sense of 'being there*" (Minsky, 1980). Heeter concluded: *A question to guide designers of virtual worlds is how do I convince participants that they and the world exist?* (Heeter, 1992).

Slater and Wilbur proposed immersion as an objectively measurable (system) characteristic and stated that presence would be "*the potential psychological and behavioral response to immersion*" (Slater and Wilbur, 1997), opening up a pathway to (technically) manipulate presence experiences. Slater later proposed two orthogonal components of presence, the place illusion (PI) and plausibility illusion (Psi) (Slater, 2009), a separation that received wide acceptance. Of late, Skarbez et al. extended on this model as depicted in **Figure 1** (Skarbez et al., 2017). They define presence as "*the perceived realness of a mediated or virtual experience.*" They further integrate additional constructs into their model, namely, copresence and social presence and specify Psi and copresence also to affect social presence. Finally, regarding the level of objectively measurable characteristics affecting the different presence components, they claim, "*that presence arises from the immersion of the system (the sensorimotor and effective valid actions it supports), the coherence of the scenario, whether the virtual experience offers company to the user, and the individual characteristics of the user.*" (96:23).

Discussion of Current Presence-Oriented XR Theories

The proposed model by Skarbez (from now on Skarbez-model) is a well-motivated extension of the older two-component model by Slater (from now on Slater-model) based on the PI and the Psi and immersion as the sole two objectively measurable (system) characteristics. Specifically, their introduction of coherence as a separate (measurable) characteristic opens up interesting perspectives. In addition, the identification of the influence of the Psi on social presence is well-motivated. The Skarbez-model also integrates various findings from the literature about the many different aspects of presence, for example, concerning social and co-presence and hence fosters the understanding of some of the primary constructs relating to the study of virtual experiences. However, we argue that there are still potential theoretical and conceptual difficulties with the Skarbez-model, some rooting back to the older Slater-model as a precursor, for example, when we make a distinct argument against the usage of the term illusion for qualia. We start the discussion with a set of questions about the Skarbez-model's propositions:

Questions About the Selection of Constructs and Their Relations

1. Why are qualia arranged hierarchically? Shall it imply the feeling of being with someone to be *less important* than the feeling of being there? Any importance does not emerge from a theoretical order but from the kind of interaction and the kind of experience per se. Other qualia, such as the virtual body ownership illusion (VBOI), seem excluded arbitrarily, despite its indicated impact on presence, see, for example, Waltemate et al., (2018).
2. Why is the Psi affecting presence and social presence but not the PI? We argue that a successful PI is affected by the coherence and plausibility of spatial cues. Hence, the Skarbez-model seems overly restrictive in its integration of plausibility in the overall theory, and, similarly, for its integration of coherence only contributing to the Psi.

Questions About Construct Layers and Construct Status

3. Why is presence in the Skarbez-model defined based on the perceived realness of an XR experience? We agree that realness in the sense of "in coherence with sensory stimuli by natural sources" plays a critical role in the sensory layer to achieve sufficient ergonomic qualities, for example, to avoid unwanted effects such as cybersickness (Stauffert et al., 2018Stauffert et al., 2020). However, on higher levels, for example, the cognitive layer, presence can be evoked *via* the PI by simple line renderings not resembling any real objects by form, color, or detail.

4. Why are the specific presence-related constructs called illusions? A quale is by definition a subjective conscious experience. From a perceptual point of view, an illusion occurs when a subjective perception lacks an objective representation. But, XR provides perceivable objective representations corresponding to subjective perceptions. In this sense, the Skarbez-model does identify presence as a quale and not an illusion but fails to do this for the contributing qualia, that is, place illusion, plausibility illusion, social presence illusion, and copresence illusion.

Some of these criticisms go far beyond a mere terminological debate and cannot be counteracted by a simple extension of the model. For example, when we talk about illusions throughout such models, then **we are conceptually manifesting the overall separation into reality and virtuality as a form of deception**. However, our models should be capable of convincingly describing where we assume the transfer from artificially generated stimuli to qualia occurs and that the effects on the users are indistinguishable from similar effects caused by natural (nonartificial) stimuli. That does not imply that people do not know that they are in an artificial environment (as in the film *The Matrix*). Phenomenological, artificial objects and environments engender a proximate stimulus representation that corresponds to subjective perception. **In addition, any subjective perception and experience, any qualia, must be assumed as real.**

Skarbez et al. (2017), Skarbez et al. (2020) also reflected on the said illusion problem. They defined a quale to focus on perceived realness in contrast to actual realness as a function of a system's ability to provide stimuli that match reality, that is, a function of immersion and coherence. They also suggested discriminating between the Place Illusion as an illusory (false) feeling of being in a remote or virtual place and placeness as a feeling of being in a real place. We argue that there is nothing like a "*false*" feeling. A quale is a subjective internal feeling which cannot be false or unreal, at least from a phenomenological point of view. The sensory stimulations giving rise to a quale can be artificial but do not render the effect "*false*" nor do they make the artificial stimulation "*false*". It is pragmatically just a distinction between the processes that generated said stimuli. Given a sufficient coherence between the quality of an artificial stimulus and the required or expected qualities as defined by our sensory, perceptual, and cognitive information processing layers, this distinction can subjectively vanish.

Similarly, regarding the second question mentioned above, one can argue that the introduced objective characteristic of coherence affecting the Psi which then affects the social presence illusion and presence but not the PI is motivated by model-specific definitions of the concepts of coherence and plausibility. If one restricts the latter ones to only impact on a cognitive level, then it is easier to argue that they do not necessarily also affect the PI. This makes the proposed model valid internally. Nevertheless, the introduction of concepts and terms to explain empirical findings should be carried out with care. One can, of course, define specific meanings to chosen terms upfront to precisely describe the intended interpretation. However, specifically with terms that have a common and widely used meaning, we would argue that it is best to stick with these definitions to strive for easy cognitive accessibility and make a model as much self-descriptive as possible. In this sense, we feel the Skarbez-model's concepts of coherence and plausibility to be partly misleading. They seem not to capture all potential applications within a presence theory and are restricted to a subset of concepts. For example, coherence of artificial visual stimuli with spatial cues expected on the sensory and perception layer can lead to a plausible evocation of spatial self-orientation and—depending on the degree of the substitution of visual stimuli from the physical environment around the user— an evocation of the feeling of "*being there*". Here, "*there*" would refer to a cognitive attribution of the sum of all spatial stimuli as belonging to an environment different from the physical environment around the user.

We honestly value the models by Skarbez et al. (2017) and by Slater (2009) and any predecessors not discussed here. Our criticisms are meant to motivate discussions and advancements in the development of theoretical models of XR experiences. From an HCI view, such models should not only generate a consistent theory of the interrelation and potential influences of important constructs, factors, and characteristics but also support guidelines for designers and developers to exploit the vast design space of XR experiences and their impact on human behaviors. This includes predictable impact paths and systematically measurable and manipulable variables (Wienrich et al., 2021) to acquire knowledge with practical impact.

BEYOND PRESENCE: CONGRUENCE AND PLAUSIBILITY

This section proposes an alternative model of XR experiences. It builds upon Skarbez et al. (2017) and Slater (2009), taking the raised criticism into account. It also shifts away from the presence, that is, the sense of "*being there*" (the PI) as the central quale to capture the many variations of XR covered by the Virtuality Continuum (VC) (Milgram and Kishino, 1994). The concept of the PI gets increasingly blurred once we move along the VC toward the nonsimulated real environment. At which point do we know that we are dealing with a Place Illusion, that is, something that is mainly caused by simulated content, and when do, we have to accept that spatial cues making us feel to be in a place are not simulated but stem from the real environment. Hence, in the wider scope of mixed and augmented reality (MR and AR), the PI becomes much less prominent.

In addition, XR technology and application development progress continuously, and its quality should likewise be evaluated. XR is already applied as a therapy system, mind, and behavior changer. Hence, we already know and accept that XR can bring users (real) experiences and cause (real) behavior. It might be comparable to the pragmatic quality in the user experience research. We presuppose that a technical device fulfills a specific function, but we are additionally interested in the hedonic, eudemonic, or social quality following the interaction with the device (Wienrich and Gramlich, 2020).

We follow the studies by Slater (2009) and Skarbez et al. (2017) and adopt *plausibility* as the first component. Valid alternatives for plausibility include *acceptance* or *suspension of disbelief* (Cruz-

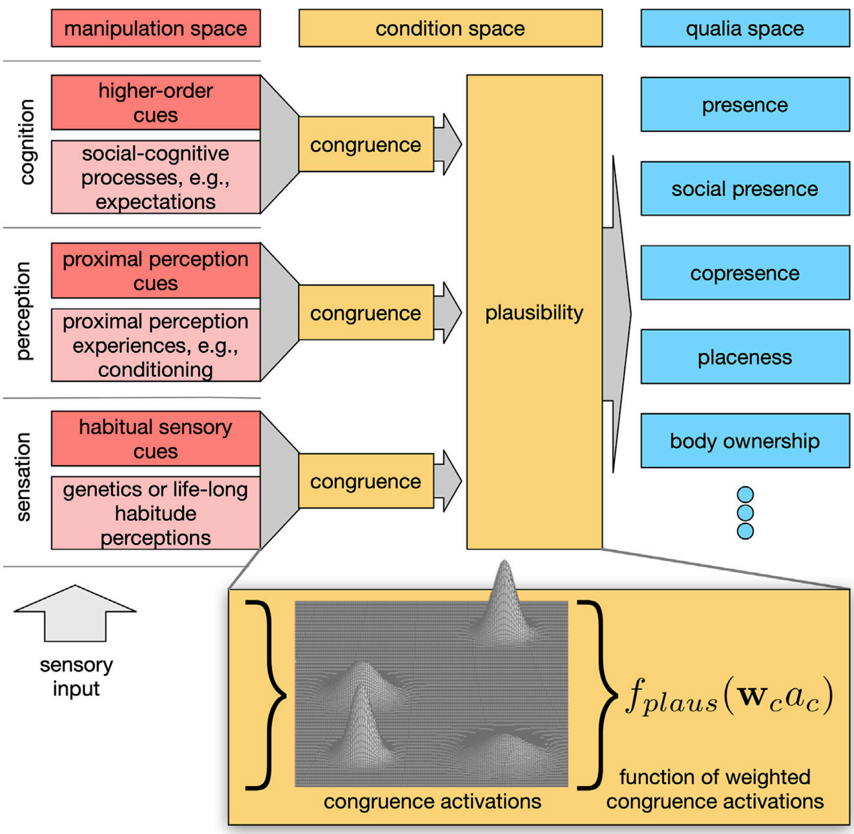

FIGURE 2 | Relationship between different XR-related qualia (including presence) and the contributing sensory information and cues proposed as an alternative new theoretical model for XR experiences and the related components. Congruence and plausibility of cues on the sensation, perception, and cognition levels take on a central role between the design and manipulation space of XR experiences and the evoked qualia. Plausibility emerges from a function of weighted congruence activations.

Neira et al., 1992; Heeter, 1992), but we focus for now on plausibility in analogy to former theoretical models. We also further specify coherence as *congruence* and include it as the second component of our proposed congruence and plausibility (CaP) model. Here, congruence is describing the objective match between processed and expected information on the sensory, perceptual, and cognitive layers.[1]

However, in contrast to the discussed presence models, we do not assume an illusion of plausibility but define plausibility as a state or condition during an XR experience that subjectively results from the evaluation and congruence of information processed by the sensory, perceptual, and cognitive layers. In our CaP-model, congruence and plausibility become central components affecting information processing on every level and giving rise to the acceptance and the suspension of disbelief (Heeter, 1992). **Figure 2** illustrates the conceptual view of the proposed CaP-model, including the main components and their relations.

The model assumes that plausibility arises from the congruence of cues on each of these layers. Each layer sets up *a frame* that defines the congruence conditions of how information is processed and interpreted and to which extent cues can be considered congruent. Here, the sensory layer exposes the base frame of information processing by setting the boundary conditions of how we transduce physical and physiological signals into neural signals. Permanently changing this frame is mainly restricted to genetic and epigenetic adaptions or cyber implants. Temporary modulation would include neuro-active drugs. The congruence conditions on this layer are accessible from biological and physiological knowledge.

In contrast to the sensory layer, the frames for the interpretation of sensory information on the perceptual and cognitive layer exhibit much more accessible plasticity and manipulation space since they are additionally also shaped by the recipient's learning, memory, knowledge, mental model, expectation, and attention, that is, proximal perception experiences and social-cognitive processes. Imagine simple animated line drawings on a 2D display. If the resulting patterns match comparable patterns generated by a perspective projection of forward/backward movements in a 3D tunnel, the

[1]Earlier versions of the CaP-model were still relying on coherence as its second component, and the first published follow-up works have adopted this. This is still valid since we see congruence as an ontological specification of coherence.

resulting perceptual congruency evokes vection independently of the underlying process generating the percepts or any degree of realness or vividness. An example of cognitive congruency is a potential appearance match of a user's avatar with her/his real physical appearance. While there is evidence that an increased match increases factors of the presence or emotional response (Waltemate et al., 2018), or acceptance (Latoschik et al., 2017), an absolute congruence is not necessary to accept the virtual body as one's own as demonstrated by the *Proteus* effect (Yee and Bailenson, 2007).

Congruency is constituted by relations between the cues and the XR experience itself. The experience can be congruent in relation to the habitual sensory cues, proximal perceptual cues, or higher-order cognitive cues. Plausibility emerges from a function of weighted congruence activations. A weighted process models dynamically changing contributions of congruent and/or incongruent relations. For example, the narrative or the use (cognitive layer) of an XR experience can be quite compelling. Then, lower sensory congruencies might contribute less strongly to plausibility and the corresponding quale. In addition, at least the sensory level of maximal congruence is reached at a certain technical advancement since user given sensory capabilities can be considered fixed. Thus, with a certain level of technological development, the level of congruence stemming from sensory relations is constantly high, but the contribution to the plausibility of emergence is still variable.

The distinction between the different sensory information processing layers allows us to pinpoint how congruence affects evaluation given a respective frame. It provides a clearer picture of the interrelated components, while it is in line with Slater's definitions of the PI to be constrained by the sensorimotor contingencies, that is, how the world is perceived and the Psi as the illusion that the scenario being depicted is occurring, that is, what is perceived (Slater, 2009). The different cue levels, reaching from bottom–up to top–down, enable prediction and empirical testing of the resulting congruence and plausibility conditions. While the bottom–up framed congruence is primarily measured objectively and quantitatively, the top–down framed congruence is mostly assessable by subjective ratings, qualitative observations, or deceiving behavioral observations. However, the suggested XR experience model allows for systematic *a priori* predictions and post-hoc explanations.

The proposed model also does not need to further define the resulting qualia's exact meaning and is largely independent of this. In other words, the model is valid for those qualia researchers, designers, and developers are interested in. The procedure to predict *a priori* or explain the post-hoc relation between the manipulated cues and the conditions experienced in XR remains the same. For example, if a definition requires us to specify a certain degree of realism (as in the Skarbez-model), then it is up to the defining instance to specify the assumed layer(s) and respective cues precisely and designers can check if they can generate such cues in congruence with the expected qualities on that layer.

DISCUSSION

This study proposed the CaP-model of XR experiences based on congruence and plausibility as central components. The proposal derived the central ideas and concepts from an analysis of promising components and potential shortcomings of existing models by Slater (2009) and later Skarbez et al. (2017). We conclude with an assessment of our model regarding important requirements (typeset in *italics*) of such a model before we discuss limitations.

In our opinion, the CaP-model possesses the *predictive and explanatory power* of modern XR experiences. The manipulation space offers realizable and systematically controllable manipulations. Well-defined frames of interpretation of the cues enable congruence checks and then *a priori* predictions or post-hoc explanations of the influence of those cues on the plausibility condition and hence the corresponding qualia. For example, if the sensory layer determines how something is sensed, objective congruence tests can assure the desired quality (e.g., to assure a required frame rate or similar technical characteristics). On higher levels, user testing might be better suited. However, despite testing for the many potential qualia, we now only have to primarily test for plausibility of the cues defined upfront as being required to evoke a certain quale.

Furthermore, our CaP-model integrated *the body of knowledge on the presence and related XR constructs*. Simultaneously, it can *avoid the aforementioned potential shortcomings of the existing model(s)*. It arranges XR experience–related qualia at one level and postulates plausibility as one common constituting and pivotal factor and a corresponding testable approach to its emergence. It shifts the focus from place illusion and is centered on three cue layers influencing congruence and plausibility and then the considered qualia. Thus, the model proposes the same prediction paths, also resolving the question of inter-qualia correlations. It resolves the often-inefficacious debate about the comparison with real-world experiences or the realness of XR experiences by accepting that XR is capable of bringing users (real) experiences and causing (real) behavior. In this sense, the proposed model identifies presence as a quale and not as an illusion and does this for other contributing qualia, such as social presence, copresence, placeness, and body ownership. Notably, we define plausibility as a *true* and for the user *real* condition during an XR experience rather than an illusion making the operationalization much easier. Questions can be formulated directly and do not rely on *as-if* comparisons.

Similarly, our CaP-model also incorporates the valid and necessary distinction between qualia and objectively measurable characteristics (Slater and Wilbur, 1997, 8), for example, as intended by the identification and definition of factors such as immersion (Slater, 1999) or company and coherence (Skarbez et al., 2017). However, our proposed model essentially simplifies such influences by identifying them as variations of just one factor our model integrates as congruence, but in a much broader context compared to (Skarbez et al., 2017), since the model incorporates congruence on all three layers of sensation, perception, and cognition.

Limitations

The present contribution is meant as a position paper taking empirical data *verifying or falsifying* the model out of scope.

However, the present study is a solid base for a set of such experiments in the future. Similarly, the *validity and soundness* requirements must be tested in future studies as well.

Finally, our proposed model simplifies complex processes as each model that tries to predict and explain human experience will have to do to a certain extent. The proposed model purposely does not claim any further details about the dependencies or interrelations between the different qualia and the resulting structure, for example, a hierarchy of factors contributing to the overall construct of presence as proposed by Skarbez et al., (2017). As we noted, in a recent experiment, we manipulated the presence and measured a correlating change in virtual body ownership, and vice versa, giving rise to speculation of an additional latent constituting factor affecting both. The latter approach highlights how these potential relationships can be investigated, and it already hints to a more complex interplay of components where the functional dependency is 1) not directed unilaterally and/or 2) hints to additional latent factors yet to be found. However, at this stage, our proposed model purposely does not try to further highlight any qualia interrelations (on the right side of **Figure 2**) since it focuses on hypothesized congruences evoking plausibility of surrounding space, embodiment, company, social interaction, and the like. Simplifications risk implicating imprecision and a lack of detail. However, they simultaneously are a necessary prerequisite for a successful generalization, which in turn helps facilitate understanding and practical usage.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

This publication was supported by the Open-Access Publication Fund of the University of Würzburg, by the XR Hub Bavaria, the German Ministry of Science and Education (BMBF) in the projects ViTraS (Grant 16SV8219) and HiAvA (Grant 16SV8781), and the EU in the project VIRTUALTIMES (Grant 824128).

REFERENCES

Cruz-Neira, C., Sandin, D. J., DeFanti, T. A., Kenyon, R. V., and Hart, J. C. (1992). The CAVE: Audio Visual Experience Automatic Virtual Environment. *Commun. ACM* 35, 64–72. doi:10.1145/129888.129892

Heeter, C. (1992). Being There: The Subjective Experience of Presence. *Presence Teleoperators Virtual Environ.* 1, 262–271. doi:10.1162/pres.1992.1.2.262

Jerome, L. W., and Jordan, P. J. (2007). Psychophysiological Perspective on Presence: The Implications of Mediated Environments on Relationships, Behavioral Health and Social Construction. *Psychol. Serv.* 4, 75–84. doi:10.1037/1541-1559.4.2.75

Latoschik, M. E., Roth, D., Gall, D., Achenbach, J., Waltemate, T., and Botsch, M. (2017). "The Effect of Avatar Realism in Immersive Social Virtual Realities", in. Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology (VRST), Gothenburg, Sweden, November 8–10, 2017, 1–39. doi:10.1145/3139131.3139156

Lombard, M., and Jones, M. T. (2015). "Defining Presence," in *Immersed in Media: Telepresence Theory, Measurement & Technology*. Editors M. Lombard, F. Biocca, J. Freeman, W. IJsselsteijn, and R. J. Schaevitz (Springer International Publishing). doi:10.1007/978-3-319-10190-3_2

Lombard, M., Biocca, F., Freeman, J., IJsselsteijn, W., and Schaevitz, R. J. (2015). in *Immersed in Media: Telepresence Theory, Measurement & Technology*. Editors M. Lombard, F. Biocca, J. Freeman, W. IJsselsteijn, and R. J. Schaevitz (Springer International Publishing). Available at: https://books.google.com/books?id=8dvyCQAAQBAJ.

Milgram, P., and Kishino, F. (1994). A Taxonomy of Mixed Reality Visual Displays. *IEICE TRANSACTIONS on Information and Systems*. 12, 1321–1329.

Minsky, M. (1980). Telepresence. *Omni* 2 (9), 45–51.

Schubert, T., and Crusius, J. (2002). "Five Theses on the Book Problem: Presence in Books, Film and VR," in PRESENCE 2002-Proceedings of the Fifth International Workshop on Presence, Porto, Portugal , October 9–11, 2002, 53–59.

Skarbez, R., Brooks, Jr., F. P., Jr., and Whitton, M. C. (2018). A Survey of Presence and Related Concepts. *ACM Comput. Surv.* 50 (96), 1–9639. doi:10.1145/3134301

Skarbez, R., Brooks, F., and Whitton, M. (2020). Immersion and Coherence: Research Agenda and Early Results. *IEEE Transactions on Visualization and Computer Graphics*, 1–11.

Slater, M., and Wilbur, S. (1997). A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. *Presence Teleoperators Virtual Environ.* 6, 603–616. doi:10.1162/pres.1997.6.6.603

Slater, M. (1999). Measuring Presence: A Response to the Witmer and Singer Presence Questionnaire. *Presence* 8, 560–565. doi:10.1162/105474699566477

Slater, M. (2009). Place Illusion and Plausibility Can Lead to Realistic Behaviour in Immersive Virtual Environments. *Phil. Trans. R. Soc. B* 364, 3549–3557. doi:10.1098/rstb.2009.0138

Stauffert, J.-P., Niebling, F., and Latoschik, M. E. (2018). "Effects of latency Jitter on Simulator Sickness in a Search Task," in *Proceedings of the 25th IEEE Virtual Reality Conference (IEEE VR)*, 121–127.

Stauffert, J.-P., Niebling, F., and Latoschik, M. E. (2020). Latency and Cybersickness: Impact, Causes, and Measures. *A Review. Frontiers in Virtual Reality* 1, 31.

Waltemate, T., Gall, D., Roth, D., Botsch, M., and Latoschik, M. E. (2018). The Impact of Avatar Personalization and Immersion on Virtual Body Ownership, Presence, and Emotional Response. *IEEE Trans. Vis. Comput. Graph.* 24, 1643–1652. doi:10.1109/tvcg.2018.2794629

Wienrich, C., and Gramlich, J. (2020). appRaiseVR - an Evaluation Framework for Immersive Experiences. *i-com* 19, 103–121. doi:10.1515/icom-2020-0008

Wienrich, C., Döllinger, N. I., and Hein, R. (2021). "Behavioral Framework of Immersive Technologies (BehaveFIT): How and Why Virtual Reality can Support Behavioral Change Processes," in *Frontiers Virtual Reality* 2, 84.

Yee, N., and Bailenson, J. (2007). The Proteus Effect: The Effect of Transformed Self-Representation on Behavior. *Hum. Comm. Res.* 33, 271–290. doi:10.1111/j.1468-2958.2007.00299.x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Virtual Reality

ORIGINAL RESEARCH
published: 23 June 2022
doi: 10.3389/frvir.2022.773448

Identifying and Coding Behavioral Indicators of Social Presence With a Social Presence Behavioral Coding System

Aleshia Taylor Hayes[1], Charles E. Hughes[2]* and Jeremy Bailenson[3]*

[1]SURGE XR Lab, Department of Learning Technologies, University of North Texas, Denton, TX, United States, [2]Department of Computer Science, Center for Research in Education Simulation Technology, University of Central Florida, Orlando, FL, United States, [3]Virtual Human Interaction Lab, Department of Communication, Stanford University, Stanford, CA, United States

Social presence, the sense of connection with another, is more important than ever as teachers, healthcare providers, and other professionals are using immersive tools to facilitate the social interaction for education, training, therapy and collaboration between geographically distributed humans and surrogates (avatars, agents, or robots). Leading researchers cite the subjective nature of the traditional self-report measures of social presence and the absence of a standardized approach to measuring social presence as a constraint to gaining deeper understanding of user's experiences of emerging and existing tools. This discourse highlights behavioral indicators of social presence that have been identified over decades across disciplines from psychology, communication, computer science, education, and engineering. The authors explicate the behavioral themes of social presence and describe a classification system grounded in exogenic and endogenic themes of social presence. This article goes on to describe the design of a social presence behavioral coding system (SPBCS) instrument that provides a structure to coding behaviors associated with a users' experience of social presence. The behavioral coding system described in this paper is the first step in creating a robust standardized approach to quantifying social presence through behavioral, physiological, and subjective indicators that ultimately may replace the current standard subjective approaches to describing the user's experience in all realities.

Keywords: mixed reality, human computer interaction, user experience evaluation, research methodology, social presence

OPEN ACCESS

Edited by:
Mary C. Whitton, University of North Carolina at Chapel Hill, United States

Reviewed by:
Ali Haskins Lisle, Virginia Tech, United States
Frank Steinicke, University of Hamburg, Germany

****Correspondence:***
Aleshia Taylor Hayes
aleshia.prof@gmail.com
Charles E. Hughes
ceh@ucf.edu

Specialty section:
This article was submitted to Technologies for VR, a section of the journal Frontiers in Virtual Reality

Received: *09 September 2021*
Accepted: *03 May 2022*
Published: *23 June 2022*

Citation:
Hayes AT, Hughes CE and Bailenson J (2022) Identifying and Coding Behavioral Indicators of Social Presence With a Social Presence Behavioral Coding System. Front. Virtual Real. 3:773448. doi: 10.3389/frvir.2022.773448

1 INTRODUCTION

The re-emergence of VR as a consumer tool, with the 2013 release of the Oculus Rift, has led to a surge in the discussions about virtual environments, beings, and the tools made available to mediate human communication in virtual spaces (Bailenson et al., 2001, 2005). In addition to the virtual collaboration tools, such as second life, that have been popular for allowing people to communicate with others while representing themselves as avatars on computer screens, new tools, such as Virbela, VRChat, AltSpaceVR, and Rec Room, afford the possibility of communicating with others in a virtual space that may be displayed in an immersive VR headset (Barreda-Ángeles and Hartmann, 2022). While it has not been established which tools are best for different use cases, it is certain that these tools still share the primary goal of allowing people to connect through computers. There are questions about the effectiveness of these computer mediated communication tools and the degree to which any of these tools can provide the experience of the face-to-face social interactions that they are meant to replace (Biocca et al., 2002; Slater, 2004; Hayes, 2015; Oh et al., 2018).

Historically, researchers and practitioners have evaluated social interaction mediated through immersive interfaces by various measures of social presence, or the sense of connection with another (Biocca et al., 2002; Slater, 2004; Hayes, 2015; Oh et al., 2018). Social presence is a sub-construct of presence derived from the constructs: telepresence and co-presence. Initially telepresence expressed the sensation teleoperators had of being at the remote worksite rather than at the operator's control station (Minsky, 1980). Over time, presence has been simplified to be "the subjective experience of being in one place or environment, even when one is physically situated in another" (Witmer and Singer, 1998, p. 1). Presence is defined as feeling as if one is "there," while co-presence describes the feeling of "being there with another". Scientists have used these constructs to evaluate the effectiveness and user experience of teleoperated, remote, and virtual experiences for decades (Chuah et al., 2013; Lu and Fan, 2016). For some experiences, e.g., operating a robot, the sense of presence is central to the effectiveness of an interaction (Lu and Fan, 2016). For other experiences, e.g., teaching a classroom of students, the sense of connection with others is critical. Further, when considering the connection experienced during a computer mediated interaction, the specific notion of presence is not always necessary. One may feel co-present with someone, even without the experience of presence, e.g. talking on the phone with someone; in contrast, one may be quite present in the physical location with another, but still feel socially disengaged, disconnected, or not socially present.

The instruments used to measure these constructs have undergone myriad changes over the years, from design and delivery to evaluation. The majority of these instruments collect subjective self-report from users of the technology. These results are useful, but often lack accuracy and users frequently misrepresent their experiences for various reasons from biases, such as a conformational bias in which the participant tries to please the researcher, to limited self-awareness. Leading researchers have even cited the futility in much of the existing design that focuses this research on subjective measures (Gunawardena and Zittle, 1997; Mennecke et al., 2011; Richardson et al., 2017). However, because of the lack of a simple substitute to the traditional subjective measure, researchers continue to use the tools we have, which are frequently subjective self-report through questionnaires, notwithstanding the acknowledged limitations. The current paper explores the development of a social presence behavioral coding system (SPBCS) meant to provide a simple quantifiable supplement to the way we understand user's experience of social presence in existing and emerging platforms.

2 IDENTIFYING SOCIAL PRESENCE THROUGH BEHAVIORS

The authors created new technique for measuring social connection experienced with a virtual other, the social presence behavioral coding system (SPBCS) to elucidate and facilitate the process of describing and quantifying the frequency of behaviors that indicate social presence. The design of this social presence behavioral coding system began with a thorough analysis of the literature and approaches to measuring social presence behaviors. The SPBCS provides specific indicators, derived from a literature review of behavioral indicators of social presence. Each such indicator was mapped to the factors of social presence. Qualitative coding is a particularly effective tool for analyzing, summarizing, distilling, and interpreting complex phenomena (Saldana, 2021). Social presence is a complex phenomenon, involving many behavioral indicators, which is why the researchers have chosen this provisional coding approach. The coding system was grounded in the literature about exogenous and endogenous behaviors that identify presence and social presence. The process was refined through the review of unstructured data to identify themes and specific codes for behaviors that indicate social presence.

This coding system is not meant to replace self-report or physiological metrics; instead, this is meant to provide a more robust approach to collecting and representing data about user behavior in regard to social presence. It is the researchers' intention that an iterative approach to testing this system will deepen and enrich the meaningfulness of the results, while providing a structured process that can be adapted to multiple researchers across contexts and domains.

2.1 Measuring Social Presence

Similar to physical presence, social presence has been measured through multiple approaches and with varying methodologies including subjective measures, behavioral measures, and physiological measures (Meehan et al., 2002; Biocca et al., 2003; Bailenson et al., 2005; Meehan et al., 2005. In addition to self-report, interviews, and various physiological measures of arousal, researchers have explored using pre-defined behavioral measures that identify a target behavior they measured as a manifestation of a user's experience of social presence; for instance, verbal and nonverbal self-disclosure has been a metric that researchers assigned as the behavioral metric by which they identified user's experience of social presence (Bailenson et al., 2005; Moustafa and Steed, 2018)). Further, researchers have explored the relationship between social presence and nonverbal synchrony, presenting the synchrony of turn taking behavior as a potential measure of social presence (Sun et al., 2019).

Throughout the literature, a consensus exists that the drawbacks of each of the approaches may be offset by integration or cross-validation with other approaches. Likewise, while the approach of measuring autonomic responses and collecting data has been effective in gaining insight into the physiological experiences that accompany the senses of presence and social presence (Cui, 2013; Meehan et al., 2002, 2005; Schlögl et al., 2002.), these criteria are hard to directly correlate to social presence as they are also used to indicate other phenomena, such as engagement, attention, and immersion (Slater et al., 2009). The literature has collectively defined several factors as either contributing to, constituting, increasing, detracting, or predicting presence (Gunawardena

TABLE 1 | Social presence factors and behavioral indicators.

Social presence factor	Behavioral indicators
Cognitive Involvement/Flow	Not noticing the time is up for session. Trying to solve problems that arise in system, attention to the needs of the virtual other
Emotional engagement (Visible display)	Laughing, smiling, nervous sweating, wringing hands, raising voice changes
Self-Disclosure	Voluntary disclosure of personal information
Valence	Intensity of emotion/Intense Emotion
Suspension of disbelief/Social Realism	Reflexive Responses: saying thank you and please, saying goodbye, trying to wrap up the lesson
Social Action/Social Actor	Respond to virtual other as if they are a social actor and not an agent
Intimacy/Immediacy	Intense emotions expressed by: raising voice, crying, laughing, orienting one's body toward the other, or other body language such as leaning towards the virtual other, or use of proximity, touch, and body orientation to indicate closeness
Physical Manipulation	Navigating the environment to "approach" other, proximity, avoidance behavior (avoiding collision)
Similarity	Reacting in ways that are consistent with similar FTF experience in this context (e.g. for teaching, trying to solve problems)
Meaningfulness of experience/Similarity	Constructing narrative of virtual other(s)/caring about them
Novelty	Expressing amazement at the technology, trying to break the system or figure out how it works
Interactivity	Balanced interplay between participant/other talk ratio
Passive Social interaction	Acknowledging the nonverbal behavior of the virtual other, mirroring nonverbal behavior of virtual other

and Zittle, 1997; Witmer and Singer, 1998; Patel et al., 2008; Thornson et al., 2009; Oh et al., 2018). **Table 1** includes those that have been identified in the literature as having a relationship to the experience of social presence. Many studies conducted over the last few decades have identified one or more behavioral indicators of social presence as their operational definition of social presence, depending on the context of the study.

2.2 Behavioral Factors of Social Presence

Early researchers outlined two inclusive subdivisions of factors of presence, exogenous or endogenous; exogenous factors are created by the generation of the virtual environment, while endogenous factors are subjective and occur within the user (Slater and Usoh, 1994; Chow, 2012). This distinction preceded a shift in research concerning presence that allowed some research to continue focusing on hardware, fidelity, and display devices, while others began to focus on the analysis of experiences in terms of the affect and cognition of the user (Parsons et al., 2009). Steuer foreshadowed this in the early writing on presence, "In other words, "presence" refers to the natural perception of an environment, and "telepresence" refers to the mediated perception of an environment" (Steuer, 1993, p. 6).

For the study of the specific phenomenon of social presence, the research is informed by the psychological study of attention in which endogenous attention refers to top-down or voluntary attention, whereas exogenous attention is considered bottom-up, or a reorientation of attention to a different stimulus of an attended stimulus (Yantis, 1993; Carretié, 2014). This classification was formerly applied to the construct of presence by early researchers of presence (Slater et al., 1994; Jin and Park, 2009). This distinction between exogenous and endogenous factors facilitates visualization of social presence factors through a structural model in which the elements interact with one another to build a deeper experience of social presence. Likewise, the distinction between exogenous and endogenous factors of social presence serve as a guide for identifying themes, categories, and codes. **Table 2** shows these factors of social presence, separated in terms of exogenous or endogenous nature.

TABLE 2 | Endogenous and exogenous factors of social presence.

Endogenous factors	Exogenous factors
Cognitive Involvement/Flow	Social Realism
Self -disclosure	Similarity
Separation anxiety	Meaningfulness of experience
Willing suspension of disbelief	Intimacy and Immediacy
Social Action/Social Actor	Physical Manipulation
Emotional engagement	Active/Passive Social interaction
Novelty	Novelty
Absence	Absence

2.3 Endogenous Factors

Endogenous factors of social presence refer to goal-driven actions directed toward the virtual other in an experience. This presents as the user consciously deciding to engage. Endogenic factors are characterized by conscious and intentional behaviors (Yantis, 1993; Carretié, 2014). Cognitive involvement, flow, self-disclosure, separation anxiety/disorientation, willing suspension of disbelief, and responding to a virtual other as a social actor constitute the endogenous factors of social presence that can be coded by an observer. Flow and novelty can be either top down or bottom up experiences and, as such, are included as both endogenous and exogenous.

2.3.1 Cognitive Involvement/Flow

Cognitive involvement is said to be an essential component to presence; similarly, it is critical to social presence (Witmer and Singer, 1998; Thornson et al., 2009). "Involvement is a psychological state experienced as a consequence of focusing one's energy and attention on a coherent set of stimuli or meaningfully related activities and events. Involvement depends on the degree of significance or meaning that the individual attaches to the stimuli, activities, or events" (Witmer and Singer, 1998, p. 227). Cognitive involvement and the state of flow are similar constructs that deal with being absorbed in an experience; and similarly, both can be either active or passive.

Flow is identified as the experience in which an individual experiences the merging of action and awareness, loss of self-consciousness, transformation of time, and enjoyment.

(Csikszentmihalyi, 1990). Flow behavior is identified by an individual's intrinsic motivation to continue, lack of self-consciousness, clarity of goals and feedback, and distorted sense of time (Csikszentmihalyi et al., 2014; Weibel et al., 2008). While flow is characterized by some of the other factors of social presence, the experience of flow can manifest itself in distinct behavioral indicators of social presence. This is related to engagement as it is "characterized by a feeling of energized focus and full involvement in the activity" (Thornson et al., 2009, p. 67). "We define passive cognitive involvement as a cognitive state in which the person is fully engaged in what s/he is doing, characterized by a feeling of energized focus and full involvement in the activity".

Cognitive involvement and flow often overlap, so, for the behavioral coding, cognitive involvement/flow is operationalized by the lack of attention to surrounding distractions, losing track of time, excitement/joy about the interaction, and expressing a desire to continue the interaction when a user is told "their time is up." Similarly, this state of cognitive involvement is indicated when a user expresses a desire to maintain the interaction or future interactions with the virtual other.

2.3.2 Self-Disclosure

While self-disclosure has been identified as an indicator of intimacy for decades (Taylor and Binder, 1973; Wheeless and Grotz, 1977), research has also maintained self-disclosure as a measure of the experience of virtual others. While researchers frequently maintain human face to face communication as the "gold standard" of interpersonal communication, some research suggests that hybrid realism possible in virtual representations may maintain high fidelity of interaction without lowering self-disclosure (Bailenson, et al., 2005).

Self-disclosure is operationally defined as an individual speaker sharing information about himself or herself that the audience (students) would not already know (Wood, 2009). Additionally, self-disclosure is a behavior that indicates intimacy and trust (Zimmer et al., 2010; Kang and Gratch, 2014). For the purposes of behavioral coding, this is coded when the participant shares personal information that the virtual other would not otherwise know. Generally, this would be unsolicited self-disclosure, as solicited self-disclosure would be compliance.

2.3.3 Separation Anxiety

The experience of disorientation or anxiety when exiting the virtual environment and returning to the physical world is the indicator of presence Witmer and Singer (1998) called Separation Anxiety/Disorientation. They indicated that the valence of this separation anxiety/disorientation might correlate with the level of presence experienced. This translates to social presence in the endogenous expression of feeling a loss or missing the agents or avatars present in the virtual space after leaving the space.

In terms of behavioral coding, this would be identified in long term interactions or in repeated interactions in which the participant identifies that they "missed" the virtual other or they seek out or request the virtual other or bring up the virtual other's characteristics as an example.

2.3.4 Suspension of Disbelief

Understanding the nature of the experiences of individuals with physical presence, co-presence, and social presence is predicated on understanding the intrinsic nature of these phenomena. Suspension of disbelief, immersive tendency, introversion, and empathy are internal experiences that relate to this research into learning within a virtual environment (Thornson et al., 2009). Originating from perceptions around media such as theater, suspension of disbelief is the phenomenon in which a participant in a virtual/synthetic/augmented environment is able to overlook and even forget the fact that the environment is not natural, but constructed and contrived, in order to enhance engagement, presence, and belief of the experience being provided/created (Boellstorff, 2011; Dede, 2009; Jeffries, 2008; Kantor, Waddington, and Osgood, 2000; LeRoy Heinrichs, Youngblood, Harter, and Dev, 2008; Maynes and et al., 1996; Park, Calvert, Brantingham, and Brantingham, 2008; Serby, 2011; Steuer, 1993).

The original concept of suspension of disbelief was actually referred to as willing suspension of disbelief, in which the implication of a conscious action on the part of the participant is central (Steuer, 1993; Serby, 2011). This idea, originating with the poet Samuel Coleridge in the early 1900s, is being challenged by the technology of the day, in which one may willingly suspend disbelief but the technology may also have the power to envelop the user yielding less power than an individual who chooses to pick up a book (Holland, 2008).

Whether it is active or passive suspension of disbelief, the suspension of disbelief is a central element to social presence. There is consistent discussion of suspension of disbelief as a contributing factor or sub-construct that can be used to describe, analyze, and measure presence and social presence (Steuer, 1993; Slater and Wilbur, 1997; Slattery, 2008). These elements of presence have been identified as creating an experience in which the technology mediating the experience "fades" or goes unnoticed or unacknowledged (Mennecke et al., 2011). Suspension of disbelief can be identified by the user ignoring the limits of the technology and behaving in ways that are consistent with believing the experience is real. For instance, if the virtual other is an agent, the user might treat them as if they are real. Likewise, if the interaction is a simulation, the use might engage within the simulation as if the experience were real and meaningful as opposed to laughing and making jokes.

2.3.5 Social Action/Social Actor

Responding to a virtual agent as if it is a social actor and not just a computer-generated object is considered a demonstration of social presence (Lombard, 2011). Slater et al. (1994) describe the phenomenon of virtual actors responding to subjects as an indicator of presence (Slater et al., 1994). This aligns with observations that participant responses to avatars in the study we present in **Section 4** may indicate levels of engagement. Their discussion was driven by Heeter's (1992) assertion that actors spontaneously reacting to the subject increases presence (Heeter, 1992). Their discussion of subjective factors further supports assumptions that an environment in which one's own body interacts with a blended physical and virtual environment, yields higher levels of presence than one where the user is being embodied by an avatar.

For behavioral coding, social action and treating the other as a social agent would be identified when participants respond to virtual agents as if they have meaningful experiences and feelings. For example, this can be seen when a user responds with interest or curiosity or even guidance when a virtual character shares a narrative. This can also be seen "when an observer treats a character in a medium as a social actor regardless of whether that actor can respond or is controlled by a human actor (e.g., watching and talking back to a TV anchor)" (Mennecke et al., 2011, p. 414).

2.3.6 Emotional Engagement

Engagement is another widely used and ambiguous term, with many denotative meanings, but the connotations of emotional engagement are essentially consistent. Engagement is generally seen in industry and in psychology as being cognitive, or focused and attentive, emotionally involved, and social, or relating to other people (Cunningham et al., 2006; Skinner et al., 2009). For the purposes of this research, engagement refers to the state of an individual being affectively, behaviorally, and cognitively involved with a virtual other in an experience. This manifests in attention, interest, and motivation to continue.

In terms of behavioral coding, this would be coded when a participant exhibits prolonged gaze/eye contact, body orientation toward the virtual other, smiling, nodding, and minimal encouragement while the virtual other is speaking. Each instance of these behaviors should be coded, for instance if someone if smiling and nodding, both behaviors should be coded.

2.4 Behavioral Indicators of Exogenous Factors

Exogenous factors of social presence are those generated by the experience in which the user has a "bottom-up" experience due to the interaction with another (Yantis, 1993; Carretié, 2014). Exogenous behaviors are the automatic behaviors that are elicited by the interaction or environment (Slater et al., 1994). The exogenous factors that can be coded through presentation of specific behaviors are social realism, similarity, attributing meaningfulness to an experience, exhibiting similar reactions in face-to-face interactions with another human, intimacy and immediacy as observed by valence of emotions expressed verbally or nonverbally, and physical manipulation.

2.4.1 Social Realism

Scholars note that the Ethopoeia phenomena (Nass and Moon, 2000) reveal that situations or social cues may trigger social action automatically. Similarly, Transformed Social Interaction (TSI) and the Proteus effect are additional theoretical frameworks through which researchers explore the reciprocal relationship between computer-mediated interaction on social behavior (Bailenson et al., 2005; Yee and Bailenson, 2007). Not only do these studies demonstrate the impact of transformed gaze and transformed proximity, they demonstrate automatic reactions, based on random distributions generated by algorithms within systems.

For behavioral coding, social realism is characterized by automatic social responses or automatic feedback responses (e.g., replying to a greeting, turn taking, lack of willingness to walk away from virtual partner when they are talking, or saying goodbye when leaving). Social realism can also be seen in the automatic responses that are characteristic of human interactions, such as turn taking or making eye contact (Ning Shen and Khalifa, 2008).

2.4.2 Similarity/Homophily

Similarity, or "homophily" refers to expressing sentiment that the other is similar in terms of attitudes or emotions. Similarity refers to the perception that the other shares attitudes, behaviors, or emotions that create a sense of social attraction.

For behavioral coding, this would be coded when a participant explicitly states their perceived similarity to the other, particularly when used in a positive tone. This may also manifest as the participant suggesting friendship or future interactions with the other. Other behaviors that would constitute similarity could be cross-coded as social realism and active passive social interaction.

2.4.3 Meaningfulness of Interaction

Meaningfulness, in relation to virtual objects and environments, refers to realistic perceptual organization (Slattery, 2008). "Meaningfulness pertains to user motivation, task saliency, and previous experience. A more meaningful situation will increase user presence" (Nam and Johnson, 2006, p. 22). While meaningfulness of experience was used in the early discourse about presence, it is appropriate to apply this factor to social presence. In fact, meaningfulness of experience was discussed prior to discussions of "social presence" in virtual experiences. The idea of social presence could easily be substituted for presence in this early analysis, "Presence should increase as the situation presented becomes more meaningful to the person. Meaningfulness is often related to many other factors, such as motivation to learn or perform, task saliency, and previous experience" (Witmer and Singer, 1998, p. 230). This is best illustrated when looking at meaningfulness as immediacy of control, authenticity of the responses and consequences of one's actions (McGreevy, 1992; Witmer and Singer, 1998).

In terms of behavioral coding, this is identified by self-disclosure of the experience as meaningful, the participant speaking about how the interaction has changed them, or them reflecting on the possibility of future interactions. Meaningfulness of interaction can also be coded when the user's disclosures demonstrate connection between their lives and the interaction or the virtual other.

2.4.4 Intimacy and Immediacy

Intimacy and immediacy refer to the intensity of emotion expressed (Biocca et al., 2003) and the nonverbal indicators of immediacy, such as proximity, touch, and body orientation

(Mehrabian, 1967). Gunawardena (1995) explains that immediacy (and non-immediacy) can be conveyed nonverbally and verbally. Research acknowledges that the intimacy and immediacy are both a cognitive state and a that is typically used to describe behaviors (Biocca et al., 2003), but computer mediated virtual interactions may be measured, controlled or mediated by the design and affordances of interfaces (Palmer 1995; Biocca et al., 2003). Because of this, intimacy and immediacy could be considered both endogenous and exogenous factors, as their manifestations are constrained by the affordances of the environment.

Intimacy and immediacy are operationalized for the behavioral coding system when the user expresses intense emotions by raising his/her voice, crying, laughing, orienting one's body toward the other, or other body language such as leaning towards the virtual during verbally expressions of intimacy. For example, leaning into the person while engaging in self-disclosure and mirroring of nonverbal behavior indicate rapport (Gratch and Lucas, 2021). Likewise, vocalic indicators, such as voice quality, pitch, rhythm, tempo, resonance, control, and accent, as well as chronemic indicators such as waiting for the other person to speak and talk time can be coded as intimacy/immediacy (Mennecke., et al., 2011). Finally, expressing desire to touch the other (e.g., I want to hug you right now) also indicate intimacy and should be coded as such.

2.4.5 Physical Manipulation

The factor of physical manipulation refers to the effort to physically manipulate objects for another in a virtual space or to ask another to manipulate physical objects, including oneself. Physical touch is associated with social presence and trust in both virtual spaces and the physical world (Chaplain, Phillips, Brown, Clanton, Stein, 2000; Oh et al., 2018).

For behavioral coding, this is operationalized by physically engaging with the virtual other. For example, accepting something that the virtual other hands to the user. Likewise, the use of proximity to engage with the virtual other is a manifestation of physical manipulation. Similarly, changing one's physical orientation to engage with the virtual other is another approach to physical manipulation, for example, sitting down next to a virtual other or walking with them. Finally, expressing desire to touch the other as a form of manipulating the interaction (e.g., I wish I could walk up behind them).

2.4.6 Active and Passive Social Interaction

Another key consideration about social presence throughout the literature is the contrast between active and passive social presence (Lombard et al., 2009). As Slater noted in reference to presence, "it is argued that reality is formed through action, rather than through mental filters" (Slater, 2004). Similarly, researchers distinguish social presence to include verbal or physical action, whereas they refer to perceiving the other as passive social presence (Lombard, 2011). Lombard. (2011) aptly describe active social presence in terms of how often a social actor engaging with an environment makes sounds out-loud, such as laughing or speaking or smiling in response to something that the other social actor does. Inversely, they describe passive social presence as observing the nonverbal behaviors of the other social actor, such as facial expression and tone.

For the social presence behavioral coding system (SPBCS), active social interaction is operationalized as the apparently automatic sounds out-loud, such as laughing or speaking or smiling in response to something that the other social actor does as well as facial expressions and tone not coded elsewhere. Likewise, passive social interaction is operationalized as automatic responses, such as frowning, grimacing, or rolling one's eyes at the virtual other, not recorded elsewhere. This could also extend to mirroring behaviors in which the participant mirrors the behaviors of the other (Ning Shen and Khalifa, 2008).

2.4.7 Absence as a Measure of Social Presence

Another effective measure of social presence was derived by Schultze in which presence is contrasted with "absence" as an occurrence in which "an individual retreats from the shared world of the here and now into a private, internal and imagined world of the mind" (Schultze and Orlikowski, 2010, p.436). The distinction between presence and absence, drawn by (Waterworth et al., 2010), refers to the attention or inattention to internal or external stimulus. Specifically, they identify absence as "psychological focus on conceptual processing," whereas, for them, presence is the psychological focus on direct perceptual processing (pg1).

For the purposes of this coding system, absence is indeed the inverse of presence, as when one becomes more present in the virtual world the loss is to presence in the physical world (absence from the physical world), and vice versa. For behavioral coding, absence would be identified by not noticing stimuli in the physical environment, due to engagement with the virtual other. Since the construct of absence is operationally defined as absence from the physical world, the coding is specifically about indicators that the user is absent from the physical world and inversely present in the "other" world. The measure of absence is the exogenic factor that relates to the endogenic factor of flow and is behaviorally indistinguishable from flow.

2.4.8 Novelty

It is difficult to extricate the impact of the novelty of experiencing virtual environments from the variables of motivation and engagement. This is particularly true of the subjective perspective of users who may experience fun but not be able to identify or pinpoint the stimulus that generates the effect of pleasure (Taylor and Binder, 1973; Slater et al., 2006; Gibson et al., 2012). "Technological novelty is the quality of perceiving digital platforms as unfamiliar, interesting, and unlike those presently used or understood" (Tokunaga, 2013, p. 3). The effect of novelty experienced can be positive or negative on learning, transfer or sense of engagement (Taylor and Binder, 1973; Jacko and Sears, 2008; Tokunaga, 2013). Four dimensions of novelty relevant to virtual interactions are thrill, change from routine, boredom alleviation, and surprise. This must be taken into consideration when evaluating both the user experience and the effectiveness of a virtual environment or experience.

For behavioral coding, novelty is operationalized by the individual explicitly makes statements of awe or wonder during the interaction (e.g. "Wow, this is so cool"). Similarly, expressions of curiosity about the system and surprise about the technology are coded as novelty.

2.4.9 Interactivity

Researchers have asserted the reciprocal nature of the relationship between interactivity and social presence (Gunawardena, 1995; Tu and McIsaac, 2002; Mykota, 2018). Heeter (1992) described interaction in a shared space with shared objects as interactivity. Tu and McIsaac (2002) discuss interactivity in terms of synchronous and asynchronous behavior, but this coding system is focusing on synchronous communication. In some instances, an individual may be eliciting interaction with the virtual other, while in other's they may be automatically responding to the virtual other. Because of this reciprocal relationship, interactivity can be either endogenous or exogenous. For the social presence behavioral coding system, interactivity refers to interplay between the user and the virtual other.

For the purposes of behavioral coding, interactivity indicators should be coded when the participant responds to the virtual other in a synchronous exchange in a way that is not coded in one of the other constructs. For this reason, this construct of interactivity should be coded to indicate interplay between interactants, such as the number of times the speaker changes (e.g., participant changing from speaker to listener every 15 s or every 3 min) (Dawson and Lignugaris/Kraft, 2017; Hayes, 2015; Li et al., 2021; Miller et al., 2021). Ideally, long term work can be done to accurately capture interplay between conversant that can be represented as a ratio of talk time and the number of times the speaker switched in an interaction.

3 METHODS

A new technique for measuring social connection that a person is experiencing with a virtual other, the social presence behavioral coding system (SPBCS) was applied to users of a mixed reality classroom. The social presence factors in the literature was used to collect frequency counts of the behaviors listed that indicate each factor represented. Each of these factors of social presence was from the literature reported in **Section 2**. The coder was instructed to place a mark in the box for each occurrence of the noted behavior during the time frame. In this case, the time frame was 8 min. There was a box available for notes to be taken for future consideration, including duration and intensity of the behavior.

3.1 Participants

The participants for this study were 22 active college instructors (10 women and 12 men) of undergraduate and graduate students at a large Southeastern university. Participants elected to complete a professional development activity to improve their teaching. These participants were screened to include only individuals who had not been exposed to the testbed, so as to eliminate familiarity with the technology as a possible confound.

FIGURE 1 | The teacher practices teaching human in the loop virtual students in the mixed reality classroom simulation training system. Teacher is observed by virtual students through a web camera and voice over internet protocol software.

3.2 Materials

3.2.1 TeachLivE Mixed Reality Classroom Simulator Test Bed

The pilot test for the social presence behavioral coding system (SPBCS) was conducted in the test-bed of TeachLivE, a Mixed Reality classroom simulator in which participants interact with a classroom of five virtual students displayed on a large screen, as shown in **Figure 1**. The virtual students in the TeachLivE environment are controlled by one human actor in the loop who is trained to simulate student behavior with higher levels of behavioral fidelity than artificial intelligence is currently capable. The user interfaced with the virtual students by delivering their lesson to the virtual kids presented on a large monitor through video conference tool. The interaction was recorded and archived in the after-action review system for review.

3.2.2 Social Presence Questionnaire as Instrumentation

The social presence instrument included in this research is the Bailenson Social Presence Instrument from a 2006 study on embodied agents designed to measure social presence. The concise nature of the instrument works well to minimize participant fatigue. Also, the items have been used in the past for a similar research approach of comparing the data from more than one method (Bailenson et al., 2005). The questions on this instrument are delivered with a Likert scale (Appendix C).

3.2.3 Validated Social Presence Self-Report Measure as Instrumentation

The social presence instrument included in this research is the Bailenson Social Presence Instrument from a 2006 study on embodied agents designed to measure social presence. The

TABLE 3 | Behavioral coding system for social presence.

Factors	Behavior	Description (Duration/Intensity)
Cognitive Involvement/Flow	Not noticing the time is up for session. Trying to solve problems that arise in system, attention	
Emotional engagement (Visible display)	Laughing, smiling, nervous, sweating, wringing hands, raising voice	
Self-Disclosure	Voluntary Disclosure of information other wouldn't know (not solicited)	
Intimacy/Immediacy	Intense emotions expressed by: raising voice, crying, laughing, orienting one's body toward the other, or other body language such as leaning towards the virtual other, or use of proximity, touch, and body orientation to indicate closeness	
Valence	Intense display of emotion (e.g. raising voice, prolonged laughing)	
Suspension of disbelief/Social Realism	Reflexive Responses: saying thank you, please, goodbye, trying to wrap up the lesson	
Social Action /Social Actor	Respond to virtual student as if they are a social actor in the world and not an agent	
Physical Manipulation	Reacted in ways that explicitly demonstrate a sense of similarity with the other	
Similarity/Homophily	Reacted in ways that are consistent with human kids (e.g. Try to solve problems, respond to questions)	
Meaningfulness of experience/Similarity	Constructing narrative of the virtual other/statements or questions that indicate caring about them	
Novelty	Expressing amazement at the technology	
Interactivity	Balanced interplay between teacher talk: student talk ratio	
Active/Passive Social interaction	Acknowledging nonverbal behavior of the students (e.g. posture, gaze, fidgeting) or mirroring nonverbal behavior of virtual other	

TABLE 4 | Revised social presence behavioral coding system (SPBCS).

Factors	Behavior	Frequency	Description (Duration/Intensity)
Cognitive Involvement/Flow	Not noticing the time is up for session. Trying to solve problems that arise in system, attention		
Emotional engagement (Visible display)	Laughing, smiling, nervous, sweating, wringing hands, raising voice		
Self-Disclosure	Voluntary Disclosure of information other wouldn't know (not solicited)		
Intimacy/Immediacy	Intense emotions expressed by: raising voice, crying, laughing, orienting one's body toward the other, or other body language such as leaning towards the virtual other, or use of proximity, touch, and body orientation to indicate closeness		
Suspension of disbelief/Social Realism	Reflexive Responses: saying thank you, please, goodbye, trying to wrap up the lesson		
Physical Manipulation	Navigating the environment to "approach" kids/ask kids to perform physical task		
Similarity/homophily	Reacted in ways that explicitly demonstrate a sense of similarity with the other		
Meaningfulness of experience/Similarity	Constructing narrative of the virtual other/statements or questions that indicate caring about them		
Interactivity	Balanced interplay between teacher talk: student talk ratio		
Active/Passive Social interaction	Acknowledging nonverbal behavior of the students (e.g. posture, gaze, fidgeting) or mirroring nonverbal behavior of virtual other		

concise nature of the instrument works well to minimize participant fatigue. Also, the items have been used in the past for a similar research approach of comparing the data from more than one method (Bailenson et al., 2004). The questions on this instrument are delivered with a Likert scale (Appendix C).

3.2.4 Social Presence Behavioral Coding System (SPBCS) as Instrumentation

This tool and training for this tool guided the rating of user behavior by two raters. Coders using the instrument were trained, using the coding sheet and review of videos, to gain a shared understanding of potential manifestations of each of the themes and coded. To save space, the coding system shown in **Table 3** has eliminated the columns for tally marks and comments that are on the actual coding system.

3.2.5 After Action Review Coding System

Each session was recorded using the After-Action Review (AAR) Coding system that was created to allow users a period of directed reflection that follows the training experience in the simulation. The AAR sessions record the interaction by recording the view of the simulation and the user. While the system is frequently used to reinforce desired behaviors and extinguish undesired behaviors, this pilot did not review the video with the participants. The videos were used for coding behaviors that indicate social presence.

3.3 Design

The dependent variables in this study are Presence and Subjective Social Presence and Observed Social Presence. Presence was measured by the modified presence questionnaire; self-reported social presence was measured with the social presence instrument and qualitative responses about the experience. Finally, behavioral social presence was measured by the behavioral social presence score from the social presence behavioral coding system (SPBCS). Consistent with the literature, the objective measure added to the study is physiological data, by way of participant heart rate (Meehan et al., 2002; Meehan et al., 2005).

3.4 Procedure

Participants were asked to engage with the students for 8 minutes, while being observed by the researcher and human in loop controlling the virtual students through the video over internet protocol software. The participants delivered their lessons while interacting with the virtual students on the 72" HD monitor, and while being observed by the researcher and recorded by the After-Action Review coding system. These videos were later reviewed for the researchers to code the participant behaviors. During the interaction, participants (faculty at a southeastern university) were told that they were part of a career day and to talk to the kids briefly about the importance of higher education. They were encouraged to participate in this experience, getting exposure to a virtual classroom and to report their perspectives. Participants were told, "keep in mind that the virtual students are able to see and hear you, but not to physically interact with you, as they are represented by avatars on a large screen but are not physically in the room." While the virtual students are all controlled by one human in the loop interactor that works to make the students display realistic behaviors of middle school students, the participants were not told if the students were agents or avatars.

3.4.1 Coding Behaviors Using the Social Presence Behavioral Coding System

The researchers compiled each of the empirical measures for social presence identified in the literature into a coding system for use while observing an interaction. The coding system instructs observers to record the frequency of the targeted behaviors as shown in **Table 3**: Cognitive Involvement/Flow, Emotional Engagement, (laughing, sweating, raising voice, raising hands), Self-Disclosure, Valence Emotion (intensity) Suspension of Disbelief Social Realism, Social Action (response to agent as if they are a social actor), Physical Manipulation, Similarity, Meaningfulness of Experience/Similarity to Real World (manifested by constructing a narrative for the student and caring about them), and Novelty (expressing amazement at the technology).

In this pilot study, two raters reviewed the video and coded each of the social presence factors exhibited by the participant (the teacher) in the virtual rehearsal. When the raters did not agree, they discussed the discrepancy until they agreed upon an answer.

3.4.2 Interpreting the Social Presence Coding System

While there were no disagreements as to whether a behavior constituted a behavioral marker of social presence, there were some discussions as to where to code a behavior. For instance, when a participant enthusiastically told the virtual students about their experience in college while laughing and smiling, one coder considered that suspension of disbelief, social realism, valance, and intimacy; while the other coder only noticed it as disclosure. This happened most frequently when deciding whether to code a behavior in the category of social actor or active social interaction or similarity. The coders had between four to seven discussions on how to code behaviors like these per participant. The approach taken in this study was to code user behaviors in as many social presence behavioral factors as they fit, which did mean that some behaviors were coded in multiple categories. Future iterations of this research will follow a more traditional approach to interrater coding to allow for double blind evaluation of interrater reliability. Likewise, future uses of this coding system will integrate some of the factors.

No new codes or coding categories emerged, instead; the coding process led the researchers to combine some categories of factors into a single category. For the coding process, the construct of emotional engagement needed to be clarified with greater specificity, as there was a great deal of overlap with the rather broad construct with other, more specifically defined constructs (e.g., active social interaction, valance, and intimacy). We decided to code actions that could be either social engagement or more specifically defined constructs as the other construct, due to the very broad nature of social engagement, and only to code social engagement if the behavior did not fit into another, more specific category. Similarly, the construct, suspension of disbelief has a great deal of overlap with social realism and social action/social actor. Because of the overlap, the other variables were collapsed into the broader construct of suspension of disbelief. This also facilitated the coding process.

3.4.3 Coding Examples

While it is important to remember the importance of utterances in the coding process, the authors have included some examples of how different physical movements should be coded, according to this first iteration of the SPBCS. Individual utterances may enhance or detract from the detection of social presence behaviors, which is to be explored in greater depth in a training video. **Figure 2** demonstrates examples of some of the physical behaviors that align with the coding standards. For **Figure 2A**., a coder should code based solely on participant utterances, as they are facing away from their conversant(s), therefore the coder can only code what the speaker says and perhaps what they write on the board. For **Figure 2B**., a coder should code emotional engagement (the teacher is smiling) and immediacy (the gaze toward the conversant), in addition to participant's utterances. A coder should code **Figure 2C** with codes of emotional engagement (smiling and hand gestures). For **Figure 2D**., a coder should code the laughter as emotional engagement and the participant's action of throwing back her head can be used as an indicator of the intensity/valence of the behavior. In **Figure 2E**., the participant is pointing to a student to call on them. This would be coded as active social interaction (pointing at the student) and passive social interaction (gaze directed toward student). The coder should code **Figure 2F** with codes of similarity and physical manipulation, due to the teacher's use of proximity to the student and approaching the student. **Figure 2G** should be coded with codes of Suspension of disbelief/Social Realism to reflect the teacher's waving goodbye to the students.

4 RESULTS

The researchers began with qualitative coding using the SPBCS grounded in the literature and social presence and moved to

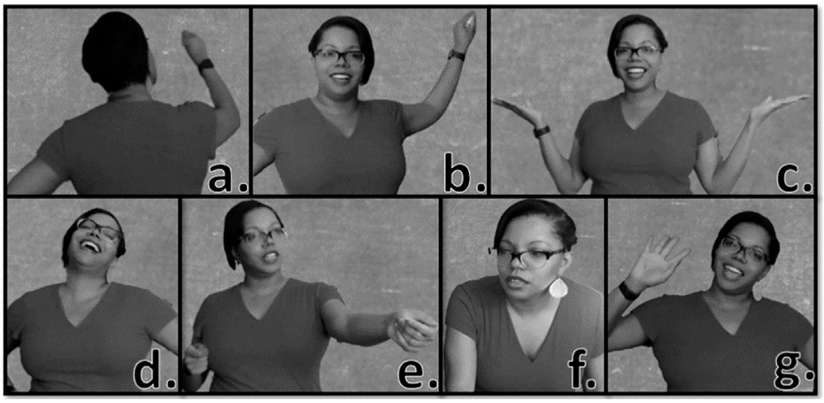

FIGURE 2 | Sample behaviors to guide SPBCS coding include **(A)** facing away from the interactant, **(B)** smiling, gaze, **(C)** gestures, **(D)** laughing, **(E)** calling on students, **(F)** using proximity, **(G)** and waving hello or goodbye. Coding behaviors and gestures should accompany coding of the participants utterances, as they frequently inform interpretation.

analysis of unstructured data to identify the themes. During the user study, we collected both qualitative subjective data and quantitative data to test and validate the Behavioral Coding System. The results are presented in the following section. Data for one participant was not included in this analysis, due to corruption of the video capture. Data for another two participants was not included, due to problems with the heart rate tracking.

4.1 Data Analysis

Each 8-min classroom simulation took approximately 1 hour to review, code, discuss and finalize the analysis. In order to come to agreements about the frequency of behaviors, the coders assigned each construct a ranking of Very High, High, Moderate, Low, or null rather than counting each instance of the behaviors aligned with each factor of social presence. This approach aligned the ordinal data from the Likert scale social presence questionnaire data.

4.1.1 Collapsing Coding Categories

The researchers removed valance and indications of absence or presence as behavioral indicators of social presence for this study, as there was a great deal of disagreement between raters on when to tag a behavior as presence or absence. Specifically, Presence/Absence was removed because the mixed reality nature of the test bed relies on cues in the virtual content that lead individuals to interact with the real and vice versa, so that can lead to an incorrect coding of absence. Finally, the researchers removed th,e ratings of novelty, as novelty of the technology could serve to distract from the user's actual sense of connection with the virtual others.

The remaining factors included: Cognitive Involvement/Flow, Emotional Engagement, Self-Disclosure, Intimacy/Immediacy, Suspension of Disbelief/Social Realism, Physical Manipulation, Similarity, Meaningfulness of Experience, Interactivity, Active and Passive Social Interaction were all included.

4.2 Data Analysis

When we classify the comprehensive Behavioral Coding System Frequencies and the Bailenson Social Presence instrument responses as Low, Very Low, Moderate, High, or Very High; all of which was in the 8-min session. Nine of the 19 participants ranked the same in the coding system as they did in the self-report, only one participant showed two levels of difference between the self-report and the Behavioral Coding System, as shown in **Figure 4**. This participant self-report indicated a moderate level of social presence, while the participant's behaviors demonstrated high social presence. Upon review of the qualitative interview responses, the participant's reflection also suggested high levels of social presence, including remembering the names of the virtual students, expressions of amazement with the technology, and the statement, "I didn't even consider that they are virtual kids (laughs) I play too many video games". Similarly, when comparing the discrepancies between the Behavioral Coding System and Social presence Questionnaire with the participant's interview reflections, the behavioral coding system ranking was more closely aligned with the reflections. Likewise, this participant was aware of the human-in the loop design of the system, which may have influenced the responses to the social presence questionnaire.

Initial analysis comparing the Social Presence Scores and the totals of the Social Presence Behavioral Coding System, using an ANOVA test, did not show any relationship between the self-report and our coding system. This led us to rethink our analysis of the data and look at the Behavioral Coding System as Ordinal Data. For this analysis, we included Cognitive Involvement/Flow, Emotional Engagement, Self-Disclosure, Intimacy/Immediacy, Suspension of Disbelief/Social Realism, Physical Manipulation, Similarity, and Meaningfulness of Experience. Interactivity proved to be another problematic construct for the strict coding, so we removed it from our analysis as well.

The scores for the Social Presence Instrument and the Behavioral Coding System rankings were analyzed using Kendall's Tau (Puka, 2011), which measures the strength of relationships between ordinal variables. Based on the results of the study, those with higher self-report of experiencing social presence were also more likely to demonstrate higher frequencies of behavioral indicators of social presence, the Kendall's Tau Phi coefficient r_t = 0.44, p <0 .05. Similarly, when a multiple

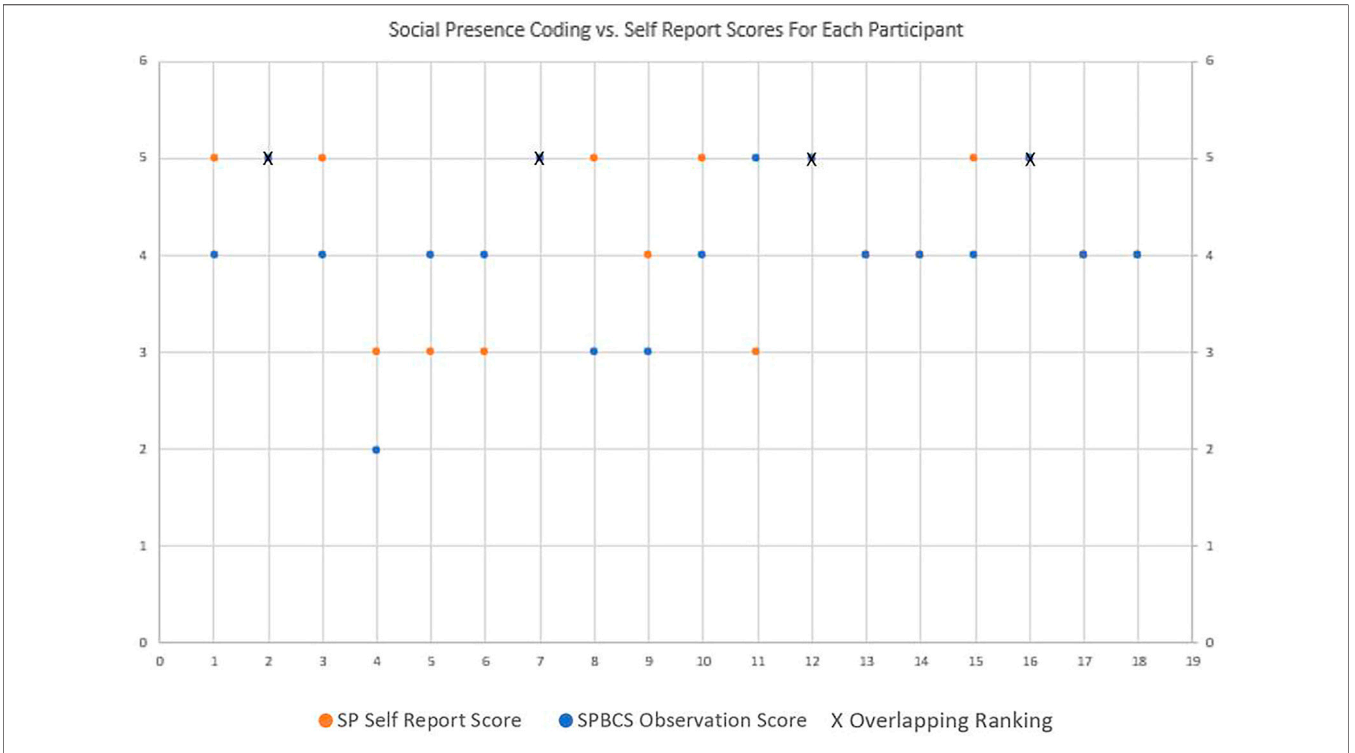

FIGURE 3 | Social Presence Coding vs Self-Report Scores demonstrate the trend of social presence coding scores largely correlating with the self report scores of social presence. This ordinal ranking.

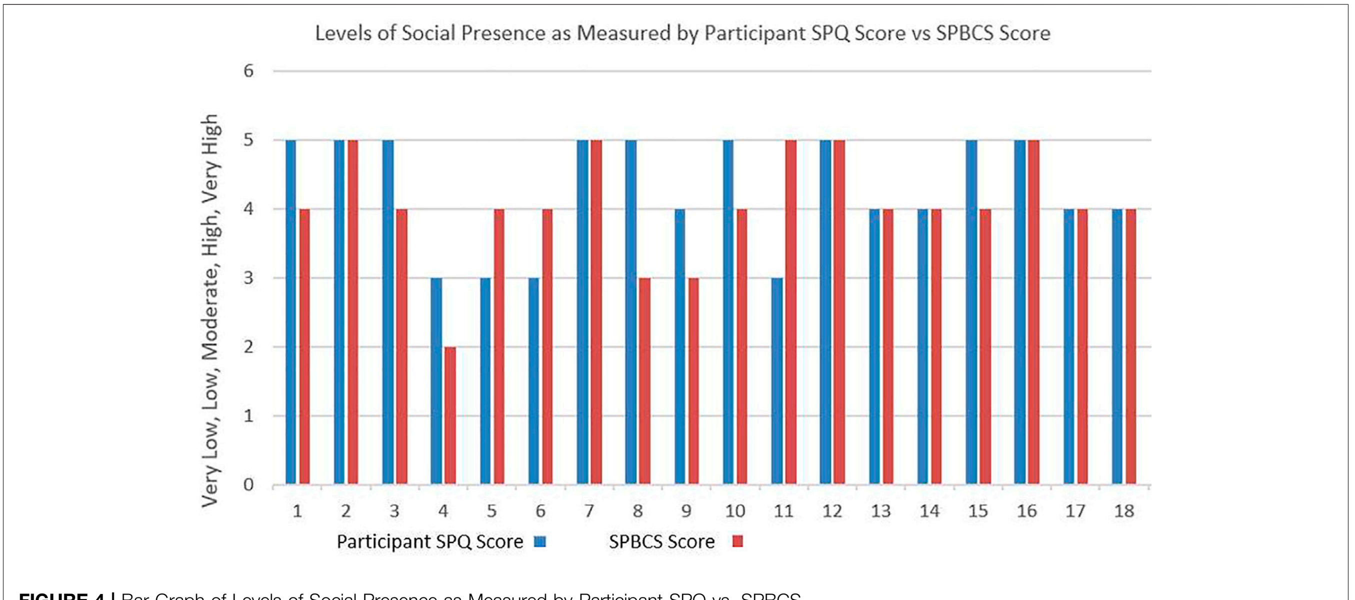

FIGURE 4 | Bar Graph of Levels of Social Presence as Measured by Participant SPQ vs. SPBCS.

regression analysis was used to test if the Behavioral Coding System could be used to predict the subjective report of the presence, the results of the regression indicated that the Behavioral Coding System significantly predicted Social Presence Composite scores (R^2 = 0.74, F (1,17) = 47.16, p <0 .01), as shown in **Figures 3**, **4**.

5 DISCUSSION

The social presence behavioral coding system (SPBCS) described in this article is the first step in refining a deductive approach to analyzing, codifying, summarizing, and interpreting human behaviors that indicate social presence within certain contexts.

Qualitative coding systems are, by nature, subject to continuous evolution, as researchers refine their understanding of phenomena and contexts and subjects are also in regular flux (Weston et al., 2001; Saldana, 2021). The initial codes that were established and refined in this study can inform the refinement of endogenic and exogenic indicators of social presence. The most promising outcome of this process has been the refinement of the codes through the collapsing of codes that overlapped. The original coding system, which included 13 codes, was refined to ten codes. This also led to a more specific and granular understanding of each code, that could only be done through trial.

5.1 Summary and Interpretation of the Results

The work to create a social presence behavioral coding system (SPBCS) tested in this pilot validation project provides critical information on how to simplify the behavioral coding system. Utilizing this coding system revealed some redundant variables and others that lacked clarity. Not only does this demonstrate a need to reduce the categories on the coding system, this also demonstrated a need for clear training for the researcher before using the social presence behavioral coding system.

The fact that the initial approach to analysis that included the numeric values of the behaviors that indicate social presence, is rather informative. Looking at these values as ordinal, categorical data aligns best with the approach of classification of behaviors. Rather than looking at a specific number of (e.g. made eye contact four times or seven times) we are looking at ordered categories (e.g. made eye contact rarely or frequently).

We maintained interactivity in the coding system, as it is an important measure of social presence, but did not include it in the data analysis, as this was high for all interactions, by the nature of the simulation. This construct would be measured best as the number of exchanges in the speaker and listener in an interaction or the percentage of time there was interactivity in an interaction. Also, this would include the percentage of time each participant (human, agent, or avatar) speaks.

The fact that the regression and Kendall's' Tau demonstrated that the Social Presence Coding system could predict the subjective self-report of social presence also provides support for the application of this tool. However, the tool does provide information that is not available in the self-report. Likewise, when the self-report and the SPBCS don't align, it may be used as an indicator that more analysis of the interaction is needed.

6 CONCLUSION AND FUTURE DIRECTIONS

The Social Presence Behavior Coding System discussed in this article is the first iteration of a system grounded in literature that addresses the need for a more standardized way to objectively measure social presence. The process of testing this system, detailed in this discourse, has led to a more streamlined version of the coding system that collapses some codes found to be redundant in their description of individual human behaviors that indicate social presence (e.g. flow and absence). Likewise, as human behavior changes with context, the coding system can be adapted to accurately predict social presence within specific contexts.

6.1 Study Limitations and Future Research

Additional work will be done to continue to refine and validate the SPBCS. This will include a Delphi study plus additional iterations with more in-depth training for coders and measurement of interrater reliability. The current study relied on the agreement of coders at the time of coding, and the lack of a measure of interrater reliability is a significant limitation to the generalizability of the social presence behavioral coding system. This study revealed the importance of a standardized training system to accompany the coding system. This training will be critical and future iterations will test the system with various levels of training among coders to determine the amount of training needed to standardize this approach.

The process of coding of user's social presence behaviors on the social presence behavioral coding system took more time and analysis than originally expected. While the a priori codes and categories based in the literature were clear, overlap between codes emerged when they were applied to actually coding participant behavior. Changes that were identified during this pilot validation process ranged from oversights, such as duplicates, to classification problems in which an activity classified as more than one construct.

Because the researchers discussed the coding system and came to an agreement before the final coding, there was no opportunity to evaluate interrater reliability. Interrater reliability will be critical to validate this coding system. While that was helpful in this first cycle of coding, testing interrater reliability of the codes is the next step. The raters should have rated separately and interrater reliability should have been calculated. The next iteration of this study will be conducted with this approach.

Future applications of the SPBCS should also be mindful in the inclusion of codes, as they are appropriate to the research context, some of the codes are specific to the context of the specific virtual classroom experience that was used for this research. For instance, Physical Manipulation in this instrument is dependent on the affordances of the virtual experience. TeachLivE afforded movement, so this would be measured more easily than it would be in an environment in which physical manipulation is not an option. This would not occur as frequently in environments that do not afford movement. What this means is that accurate choice of the changes in affordance has the potential to change the score. The researchers would also suggest establishing a baseline of expected or target behaviors for each study or context. For example, the expectations for a peer to peer interaction could be measured by the SPBCS, but the baseline expectation would differ greatly from the baseline of a teacher to classroom interaction.

Future applications of this preliminary SPBCS will implement a training job aid to be reviewed by the research team before using the social presence behavioral coding sheet. This will address some of the ambiguity with the coding system. Future uses will integrate training that includes a plan to avoid classification problems in which an activity classified as more than one construct. Future training should also include the distinction of whether the team intends to code certain events as one or multiple factors. To standardize this training process, the research team will use videos of interactions from multiple contexts and code them as the training set for other researchers who want to use the instrument. All videos will be coded independently by coders from the same cultural background, the partner's cultural background, and a third, unrelated cultural background.

The next steps in the iteration on the social presence behavioral coding system will be a Delphi Study. After the necessary changes, revealed by this pilot have been addressed, the researchers will retain a panel of experts to dive into the strengths and limitations of the social presence behavioral coding system (SPBCS) to arrive at a consensus over the representation of the factors of social presence on the coding system. The most up to date version of the SPBCS can be seen in **Table 4**.

Future studies to build and validate the Behavioral Coding System should include an additional physiological metric, in order to add clarity to the ambiguity in the interpretation of heart rate data. Heart rate data, while correlated with experience of presence, can also indicate many other phenomena in the user (e.g., stress to movement). For example, spikes in heart rate when students asked about the participant could indicate that the user experienced social presence or engagement, but it could also reflect stress related to the user's personality type (introversion vs. extroversion).

Upon completion of the Delphi study and validation study, the next steps for this research will be to automate the capture of behaviors that indicate social presence. Face tracking, body tracking, and natural language processing can be used to automate and refine the process described for the Social Presence Behavioral Coding System.

Social presence is a complex human experience. Not only can this SPBCS contribute to the measurement of the degree of social presence that an individual experiences, it can also provide a deeper description of how an individual is experiencing social presence. While the researchers intend to continue to iterate on the Social Presence Behavioral Coding System, this work is a starting point for this team and other researchers to begin streamlined approaches to objectively measure and describe users' behaviors that indicate social presence. Not only will this improve the meaningfulness of research outcomes, it can guide future design. This work can also lead to a more unified theory of what to look for when measuring human connection in physical, hybrid, or virtual spaces. The work is also useful as we move to build agents and avatars to address human needs for learning, training, socialization, and interaction. The work that comes from the applications of this social presence behavioral coding system (SPBCS) can inform the development, iteration, and evaluation of agents and systems.

DATA AVAILABILITY STATEMENT

Raw data were generated at University of Central Florida. Derived data supporting the findings of this study are available from the corresponding author ATH on request.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of Central Florida Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

ACKNOWLEDGMENTS

We would like to acknowledge the principal investigators (PIs) of TLE TeachLivE, Lisa Dieker, and Michael Hynes who were not among the authors of this paper. We would like to acknowledge the Bill & Melinda Gates Foundation, whose gift in part supported the TeachLivE team and the execution of this study, and the National Science Foundation (CNS1051067) whose support contributed to the technical infrastructure.

REFERENCES

Bailenson, J. N., Beall, A. C., Loomis, J., Blascovich, J., and Turk, M. (2005). Transformed Social Interaction, Augmented Gaze, and Social Influence in Immersive Virtual Environments. *Hum. Comm. Res.* 31 (4), 511–537. doi:10.1111/j.1468-2958.2005.tb00881.x

Bailenson, J. N., Blascovich, J., Beall, A. C., and Loomis, J. M. (2001). Equilibrium Theory Revisited: Mutual Gaze and Personal Space in Virtual Environments. *Presence Teleoperators Virtual Environ.* 10 (6), 583–598. doi:10.1162/105474601753272844

Barreda-Ángeles, M., and Hartmann, T. (2022). Psychological Benefits of Using Social Virtual Reality Platforms during the Covid-19 Pandemic: The Role of Social and Spatial Presence. *Comput. Hum. Behav.* 127, 107047. doi:10.1016/j.chb.2021.107047

Biocca, F., Harms, C., and Burgoon, J. K. (2003). Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria. *Presence Teleoperators Virtual Environ.* 12 (5), 456–480. doi:10.1162/105474603322761270

Biocca, F., Harms, C., and Burgoon, J. K. (2002). Criteria for a theory and measure of social presence. Paper under review for Presence.

Carretié, L. (2014). Exogenous (Automatic) Attention to Emotional Stimuli: A Review. *Cogn. Affect Behav. Neurosci.* 14 (4), 1228–1258. doi:10.3758/s13415-014-0270-2

Chow, M. C. M. (2012). "Factors Influencing Presence in Virtual Worlds," in *NI 2012 : 11th International Congress on Nursing Informatics* (Montreal, Canada), 083. *June 23-27, 2012*.

Chuah, J. H., Robb, A., White, C., Wendling, A., Lampotang, S., Kopper, R., et al. (2013). Exploring Agent Physicality and Social Presence for Medical Team Training. *Presence Teleoperators Virtual Environ.* 22 (2), 141–170. doi:10.1162/PRES_a_00145

Cui, G. (2013). Evaluating Online Social Presence: An Overview of Social Presence Assessment. *Jetde* 6 (1). doi:10.18785/jetde.0601.02

Cunningham, T., Hall, A. S., and Young, C. (2006). The Advertising Magnifier Effect: An MTV Study. *Jar* 46 (4), 369–380. doi:10.2501/S0021849906060454

Dawson, M. R., and Lignugaris/Kraft, B. (2017). Meaningful Practice. *Teach. Educ. Special Educ.* 40 (1), 26–50. doi:10.1177/0888406416664184

Gibson, D., Osterweil, S., Zintgraff, C., Fisser, P., Wood, T., Knezek, G., et al. (2012). "Next Generation Learning Challenges EDUCAUSE WInners," in *Proceedings of SITE 2012—Society for Information Technology and Teacher Education International Conference*. Editor P. Resta (Austin, Texas, USA: Association for the Advancement of Computing in Education (AACE)).

Gunawardena, C. N. (1995). Social Presence Theory and Implications for Interaction and Collaborative Learning in Computer Conferences. *Int. J. Educ. Telecommun.* 1 (2), 147

Gunawardena, C. N., and Zittle, F. J. (1997). Social Presence as a Predictor of Satisfaction within a Computer-mediated Conferencing Environment. *Am. J. Distance Educ.* 11 (3), 8–26. doi:10.1080/08923649709526970

Hayes, A. (2015). *The Experience of Physical and Social Presence in a Virtual Learning Environment as Impacted by the Affordance of Movement Enabled by Motion Tracking*. USA: University of Central Florida. Doctoral dissertation. https://stars.library.ucf.edu/etd/5014.

Heeter, C. (1992). Being There: The Subjective Experience of Presence. *Presence Teleoperators Virtual Environ.* 1, 262–271. doi:10.1162/pres.1992.1.2.262

Holland, N. N. (2008). Spider-Man? Sure! the Neuroscience of Suspending Disbelief. *Interdiscip. Sci. Rev.* 33 (4), 312–320. doi:10.1179/174327908X392870

Jacko, J. A., and Sears, A. (2008). *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications*. nlebk: Lawrence Erlbaum Associates. Available at: http://ezproxy.lib.ucf.edu/login?URL=http://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=211061&site=ehost-live.

Jin, S.-A. A., and Park, N. (2009). Parasocial Interaction with My Avatar: Effects of Interdependent Self-Construal and the Mediating Role of Self-Presence in an Avatar-Based Console Game, Wii. *CyberPsychology Behav.* 12 (6), 723–727. doi:10.1089/cpb.2008.0289

Kang, S.-H., and Gratch, J. (2014). Exploring Users' Social Responses to Computer Counseling Interviewers' Behavior. *Comput. Hum. Behav.* 34, 120–130. doi:10.1016/j.chb.2014.01.006

Li, J. V., Fernandes, S., McVeigh-Schultz, J., and Isbister, K. (2021). "Challenges in Running Social VR Studies A Tale of a Covid Procedural Pivot," in Conference on Human Factors in Computing Systems At: Yokohama, Japan. (ResearchGate).

Lombard, M., Ditton, T. B., and Weinstein, L. (2009). Measuring Presence: The Temple Presence Inventory." in Proceedings of the 12th Annual International Workshop on Presence

Lombard, M. (2011). "Why Telepresence Matters: A Personal Answer To the "So what?"," in Presented at the 2011 Annual Conference of the International Society for Presence Research (ISPR), Edinburgh, Scotland, 2011, October.

Lu, B., Fan, W., and Zhou, M. (2016). Social Presence, Trust, and Social Commerce Purchase Intention: An Empirical Research. *Comput. Hum. Behav.* 56, 225–237. doi:10.1016/j.chb.2015.11.057

McGreevy, M. W. (1992). The Presence of Field Geologists in Mars-Like Terrain. *Presence Teleoperators Virtual Environ.* 1 (4), 375–403. doi:10.1162/pres.1992.1.4.375

Meehan, M., Insko, B., Whitton, M., and Brooks, F. P. (2002). Physiological Measures of Presence in Stressful Virtual Environments. *ACM Trans. Graph.* 21 (3), 645–652. doi:10.1145/566654.566630

Meehan, M., Razzaque, S., Insko, B., Whitton, M., and Brooks, F. P. (2005). Review of Four Studies on the Use of Physiological Reaction as a Measure of Presence in StressfulVirtual Environments. *Appl. Psychophysiol. Biofeedback* 30 (3), 239–258. doi:10.1007/s10484-005-6381-3

Mehbabian, A. (1967). Orientation Behaviors and Nonverbal Attitude Communication1. *J. Commun.* 17 (4), 324–332. doi:10.1111/j.1460-2466.1967.tb01190.x

Mennecke, B. E., Triplett, J. L., Hassall, L. M., Conde, Z. J., and Heer, R. (2011). An Examination of a Theory of Embodied Social Presence in Virtual Worlds*. *Decis. Sci.* 42 (2), 413–450. doi:10.1111/j.1540-5915.2011.00317.x

Miller, M. R., Sonalkar, N., Mabogunje, A., Leifer, L., and Bailenson, J. (2021). Synchrony within Triads Using Virtual Reality. *Proc. ACM Hum.-Comput. Interact.* 5 (CSCW2), 4001–40027. doi:10.1145/3479544

Minsky, M. (1980). *Telepresence*.

Moustafa, F., and Steed, A. (2018). A Longitudinal Study of Small Group Interaction in Social Virtual Reality." in Proceedings of the 24th ACM Symposium on Virtual Reality Softw.are and Technology, 1–10. doi:10.1145/3281505.3281527

Mykota, D. (2018). The Effective Affect: A Scoping Review of Social Presence. *Int. J. E-Learning Distance Educ.* 33 (2), 1–30.

Nam, C., and Johnson, S. (2006). *Presence in Technologically Mediated Environments: A Research Framework*.

Nass, C., and Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers. *J. Soc. Isssues* 56 (1), 81–103. doi:10.1111/0022-4537.00153

Ning Shen, K., and Khalifa, M. (2008). Exploring Multidimensional Conceptualization of Social Presence in the Context of Online Communities. *Int. J. Human-Computer Interact.* 24 (7), 722–748. doi:10.1080/10447310802335789

Oh, C. S., Bailenson, J. N., and Welch, G. F. (2018). A Systematic Review of Social Presence: Definition, Antecedents, and Implications. *Front. Robot. AI* 5, 5. doi:10.3389/frobt.2018.00114

Parsons, T. D., Kenny, P., Cosand, L., Iyer, A., Courtney, C., and Rizzo, A. A. (2009). A Virtual Human Agent for Assessing Bias in Novice Therapists. *Stud. health tech. Info.* 142, 253–258. doi:10.3233/978-1-58603-964-6-253

Patel, K., Bailenson, J., Hack-Jung, S., Diankov, R., and Bajcsy, R. (2008). *R.: The Effects of Fully Immersive Virtual Reality on the Learning of Physical Tasks*.

Puka, L. (2011). *Kendall's Tau*, 713–715. doi:10.1007/978-3-642-04898-2_324

Richardson, J. C., Maeda, Y., Lv, J., and Caskurlu, S. (2017). Social Presence in Relation to Students' Satisfaction and Learning in the Online Environment: A Meta-Analysis. *Comput. Hum. Behav.* 71, 402–417. doi:10.1016/j.chb.2017.02.001

Saldana, J. (2021). *Analyzing and Interpreting Qualitative Research: After the Interview*. SAGE Publications.

Schlögl, A., Slater, M., and Pfurtscheller, G. (2002). *Presence Research and EEG*, 7.

Schultze, U., and Orlikowski, W. J. (2010). Research Commentary-Virtual Worlds: A Performative Perspective on Globally Distributed, Immersive Work. *Inf. Syst. Res.* 21 (4), 810–821. doi:10.1287/isre.1100.0321

Serby, T. (2011). Willing Suspension of Disbelief: A Study in Online Learning through Simulation, and its Potential for Deeper Learning in Higher Education. *Liverp. Law Rev.* 32 (2), 181–195. doi:10.1007/s10991-011-9095-z

Skinner, E. A., Kindermann, T. A., and Furrer, C. J. (2009). A Motivational Perspective on Engagement and Disaffection Conceptualization and Assessment of Children's Behavioral and Emotional Participation in Academic Activities in the Classroom. *Educ. Psychol. Meas.* 69 (3), 493–525. doi:10.1177/0013164408323233

Slater, M., Guger, C., Edlinger, G., Leeb, R., Pfurtscheller, G., Antley, A., et al. (2006). Analysis of Physiological Responses to a Social Situation in an Immersive Virtual Environment. *Presence Teleoperators Virtual Environ.* 15 (5), 553–569. doi:10.1162/pres.15.5.553

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

Slater, M., Lotto, B., Arnold, M. M., and Sánchez-Vives, M. V. (2009). How We Experience Immersive Virtual Environments: The Concept of Presence and its Measurement. Available at: http://diposit.ub.edu/dspace/handle/2445/49643.

Slater, M., Usoh, M., and Steed, A. (1994). Depth of Presence in Virtual Environments. *Presence Teleoperators Virtual Environ.* 3, 130–144. doi:10.1162/pres.1994.3.2.130

Slater, M., and Wilbur, S. (1997). A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. *Presence Teleoperators Virtual Environ.* 6, 603–616. doi:10.1162/pres.1997.6.6.603

Slattery, D. R. (2008). VR and Hallucination: A Technoetic Perspective. *Technoetic arts* 6 (1), 3–18. doi:10.1386/tear.6.1.3_1

Steuer, J. (1993). *Defining Virtual Reality: Dimensions Determining Telepresence*, 25.

Sun, Y., Shaikh, O., and Won, A. S. (2019). Nonverbal Synchrony in Virtual Reality. *PLOS ONE* 14 (9), e0221803. doi:10.1371/journal.pone.0221803

Taylor, D. A., and Binder, A. (1973). Degree of Training and the Novelty Transfer Effect. *Mem. Cognition* 1 (1), 61–63. doi:10.3758/BF03198069

Thornson, C. A., Goldiez, B. F., and Le, H. (2009). Predicting Presence: Constructing the Tendency Toward Presence Inventory. *Int. J. Human-Computer Stud.* 67, 62–78. doi:10.1016/j.ijhcs.2008.08.006

Tokunaga, R. S. (2013). Engagement with Novel Virtual Environments: The Role of Perceived Novelty and Flow in the Development of the Deficient Self-Regulation of Internet Use and Media Habits. *Hum. Commun. Res.* 39 (3), 365–393. doi:10.1111/hcre.12008

Tu, C.-H., and McIsaac, M. (2002). The Relationship of Social Presence and Interaction in Online Classes. *Am. J. Distance Educ.* 16 (3), 131–150. doi:10.1207/S15389286AJDE1603_2

Waterworth, J., Waterworth, E., Mantovani, F., and Riva, G. (2010). On Feeling (The) Present:An Evolutionary Account of the Sense of Presence in Physical and Electronically-Mediated Environments. *J. Conscious. Stud.* 17, 1–19.

Weston, C., Gandell, T., Beauchamp, J., McAlpine, L., Wiseman, C., and Beauchamp, C. (2001). Analyzing Interview Data: The Development and Evolution of a Coding System. *Qual. Sociol.* 24 (3), 381–400. doi:10.1023/A:1010690908200

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

Wood, J. T. (2009). *Communication in Our Lives*. 5th ed. Wadsworth Cengage Learning, 15047060.

Yantis, S. (1993). Stimulus-driven Attentional Capture and Attentional Control Settings. *J. Exp. Psychol. Hum. Percept. Perform.* 19 (3), 676–681. doi:10.1037/0096-1523.19.3.676

Yee, N., and Bailenson, J. (2007). The Proteus Effect: The Effect of Transformed Self-Representation on Behavior. *Hum. Comm. Res.* 33 (3), 271–290. doi:10.1111/j.1468-2958.2007.00299.x

Zimmer, J. C., Arsal, R. E., Al-Marzouq, M., and Grover, V. (2010). Investigating Online Information Disclosure: Effects of Information Relevance, Trust and Risk. *Inf. Manag.* 47 (2), 115–123. doi:10.1016/j.im.2009.12.003

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership