



OPEN ACCESS

EDITED BY

Abdeldjalil Naceri,
Technical University of Munich, Germany

REVIEWED BY

Giacinto Barresi,
Italian Institute of Technology (IIT), Italy
David Murphy,
University College Cork, Ireland

*CORRESPONDENCE

Gary Bente,
✉ gabente@msu.edu

SPECIALTY SECTION

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

RECEIVED 30 August 2022

ACCEPTED 27 March 2023

PUBLISHED 13 April 2023

CITATION

Bente G, Schmäzle R, Jahn NT and Schaaf A (2023), Measuring the effects of co-location on emotion perception in shared virtual environments: An ecological perspective.
Front. Virtual Real. 4:1032510.
doi: 10.3389/frvir.2023.1032510

COPYRIGHT

© 2023 Bente, Schmäzle, Jahn and Schaaf. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Measuring the effects of co-location on emotion perception in shared virtual environments: An ecological perspective

Gary Bente*, Ralf Schmäzle, Nolan T. Jahn and Andrea Schaaf

Department of Communication, Michigan State University, East Lansing, MI, United States

Inferring emotions from others' non-verbal behavior is a pervasive and fundamental task in social interactions. Typically, real-life encounters imply the co-location of interactants, i.e., their embodiment within a shared spatial-temporal continuum in which the trajectories of the interaction partner's Expressive Body Movement (EBM) create mutual social affordances. Shared Virtual Environments (SVEs) and Virtual Characters (VCs) are increasingly used to study social perception, allowing to reconcile experimental stimulus control with ecological validity. However, it remains unclear whether display modalities that enable co-presence have an impact on observers responses to VCs' expressive behaviors. Drawing upon ecological approaches to social perception, we reasoned that sharing the space with a VC should amplify affordances as compared to a screen display, and consequently alter observers' perceptions of EBM in terms of judgment certainty, hit rates, perceived expressive qualities (arousal and valence), and resulting approach and avoidance tendencies. In a between-subject design, we compared the perception of 54 10-s animations of VCs performing three daily activities (painting, mopping, sanding) in three emotional states (angry, happy, sad)—either displayed in 3D as a co-located VC moving in shared space, or as a 2D replay on a screen that was also placed in the SVEs. Results confirm the effective experimental control of the variable of interest, showing that perceived co-presence was significantly affected by the display modality, while perceived realism and immersion showed no difference. Spatial presence and social presence showed marginal effects. Results suggest that the display modality had a minimal effect on emotion perception. A weak effect was found for the expression "happy," for which unbiased hit rates were higher in the 3D condition. Importantly, low hit rates were observed for all three emotion categories. However, observers judgments significantly correlated for category assignment and across all rating dimensions, indicating universal decoding principles. While category assignment was erroneous, though, ratings of valence and arousal were consistent with expectations derived from emotion theory. The study demonstrates the value of animated VCs in emotion perception studies and raises new questions regarding the validity of category-based emotion recognition measures.

KEYWORDS

virtual environment, emotion perception, virtual reality, co-presence, avatars, non-verbal communication

Introduction

Taking an ecological perspective on social perception (McArthur and Baron, 1983; Zebrowitz, 2002), the current study examines how variations in spatial co-location influence emotion inferences from expressive movement behavior shown by a Virtual Character (VC) within a Shared Virtual Environment (SVE). Addressing this topic is crucial for two reasons. First, recent research has increasingly used character animation as a tool to study communication and social perception, exploiting its unique capacity to unite experimental control and stimulus realism (Nowak and Fox, 2018; Pan and Hamilton, 2018; Bente, 2019). While so far, most studies used non-immersive 2D stimuli that display VC's motion on a screen, current immersive 3D display technologies allow for the presentation of character animations in a shared virtual space, which can be entered and navigated by the observer. In such a setting, similar to face-to-face encounters, the observer is directly exposed to the trajectories of a VC's expressive motions, for instance when facing a rapid or fierce movement toward her/himself. In an early study, Bailenson, Blascovich, Beall, and Loomis (2003) demonstrated that sharing space with an avatar influences locomotion and distancing behaviors relative to the virtual other, similar to real-life encounters. Still, it remains an open question whether the co-location of an observer and a VC in a shared space impacts observers' perception of expressive behaviors, as compared to a VC shown on a screen as in most of the previous studies using character animations as stimuli.

Second, *Social VR* is expected to conquer the mass market and become a core communication technology in the near future (see Bailenson, 2018). This renders the dynamics of social perception within SVEs an important research topic in itself (Fabri, Moore, and Hobbs, 2004; Blascovich and Bailenson, 2011; Sun and Won, 2021; Hepperle, Purps, Deuchler, and Wölfel, 2022; Rogers et al., 2022). While various studies have shown that the non-verbal behavior of VCs, including movements, postures, gaze, and facial expressions produce similar effects as real-life or videotaped human behavior (Bente, et al., 2001; Noël, Dumoulin, and Lindgaard, 2009; Basori and Ali, 2013), surprisingly little is known about the specific impact of immersive VR display technology on social perception. Specifically, we do not know whether the subtle emotional nuances in non-verbal communication can be more effectively conveyed in VR if observers and virtual others are co-located in an immersive SVE compared to a desktop VR setting (see Bente et al., 2008). The current study examines this question by focusing on observers' perception of VC's expressive body movements.

Background

Social affordances and the role of co-location

The ecological perspective on social perception asks which affordances emanate from the "physical qualities" of others,

including static features, such as size, weight, and attractiveness, but also, more importantly, dynamic features, such as face and body movements. The concept of affordances was introduced by Gibson (1979) in his *Ecological Approach to Visual Perception* in order to characterize the relation of perceiver and percept in terms of options or urges for action, suggesting that "perception is for doing" (see Zebrowitz, 2002, p. 143). In Gibson's terms an affordance is a stable property of the percept (the object or subject we face). Dings (2018) concludes: "The affordance of an object thus does not change, as the need of the observer changes" (p. 683). What might change dynamically though, is the valence of the affordance, i.e., *how* we perceive the affordance, and which specific actions are solicited. McArthur and Baron (1983) applied this perspective to social perception, directing researcher's attention away from a traditional 'mind reading' approach towards the adaptive function of person and emotion perception. Zebrowitz and Collins (1997) define 'social affordance' as "...an emergent property, reflecting task-relevant properties that the target person has for the perceiver" (p.218). Expressive behaviors in this sense are conceived as dynamic characteristics of other social entities (real or virtual) that afford specific actions of an observer directed towards the given environment, including the target person.

Because in this perspective perception and action are inextricably interlinked, the co-location of observer and target in the same spatial-temporal continuum is assumed to be crucial for processing expressive behaviors. On-screen presentations of social stimuli, as traditionally used in emotion perception studies, might thus fail to create relevant affordances. Immersive Virtual Environment Technologies (IVET) have long been propagated as an efficient research tool that allows us to overcome these limitations, and to enable the generation of situated, fully embodied social stimuli. Zebrowitz (2002) posits: "IVET can enable researchers to investigate social perception in a manner that fulfills many of the assumptions of the ecological approach" (p.143). Surprisingly, the use of VR technologies in social perception studies has been limited to character animations mostly displayed on 2D screens and rarely *via* immersive display technologies that allow for a co-location of observer and target in a shared space. From an ecological perspective, however, these display modalities should make a difference, as the mere possibility to move towards or away from a social target in a shared space can be considered a social affordance feature that putatively impacts social perception. This simple, but potentially consequential assumption has not been put to test yet. This is the major goal of the current study.

Dimensions of presence

Grabarczyk and Pokropski (2016) argue that affordances and embodiment are crucial factors in the experience of immersion and presence in virtual environments. Most commonly, *immersion* is used to describe a technology feature, i.e., its capacity to absorb the users' senses and focus attention on the virtual world, while the term *presence* is used to characterize a user experience in VR (see, Oh,

Bailenson, and Welch, 2018; Slater, 1999). Similar to *telepresence* (Minsky, 1980), *presence* has been defined as the “sense of being there” (Sheridan, 1992), i.e., the “subjective experience of being in an environment, even when one is physically situated in another” (Witmer and Singer, 1998, p. 225). Along with the development of VR technologies, and particularly its growing social use, the concept of presence has been progressively diversified, particularly in order to differentiate between the experience of physical and social elements in virtual environments (Lee, 2004; Bulu, 2012). This has led to a series of sub-constructs, such as spatial or place presence on the one hand (Bulu, 2012; Hartmann, et al., 2016) and co-presence and social presence on the other hand (Nowak, 2001; Oh, Bailenson, and Welch, 2018).

There is no room here to discuss the distinctions between the various facets of presence in greater detail (critical overviews can be found in Felton and Jackson, 2022; Lee, 2004; Oh et al., 2018). In fact, the boundaries between the various concepts are blurred, their usage is inconsistent, and data regarding their mutual interdependencies are equivocal (see Bulu, 2012; Cummings and Bailenson, 2016). For instance, co-presence has been defined in analogy to presence as “the sense of being there together” (Schroeder, 2006) which puts the construct closer to the construct of spatial presence. Others define “co-presence” with reference to its original use (see Goffman, 1963) as a psychological connection of minds while conceiving “social presence” as a property of a medium (Nowak, 2001). Given the variety of definitions and respective measures it isn’t surprising that data regarding the interdependencies of the different constructs is also inconclusive. While some studies found spatial/place presence and co-presence to be correlated (Axelsson, et al., 2001; Schroeder, 2002) other studies failed to corroborate this result (Bystrom and Barfield, 1999; Casanueva, 2001) or found spatial presence to be more closely related to social presence (Thie and van Wijk, 1998). In so far as there is no substantial ground truth and the different presence concepts are operationally defined through (varying) measurement instruments, it seems idle at this point to continue this discussion. Instead, we suggest a pragmatic approach to the definition problem that allows us to identify distinct perceptual phenomena relevant to the ecological approach toward social perception. With this intention, we particularly aim to differentiate between: 1) Spatial presence, i.e., the sense of being transported to another place, 2) co-presence, i.e., the sense of being co-located in that place with another actor, and 3) social presence, i.e., the sense of being psychologically connected in terms of attention and emotional contagion (see Bente et al., 2008). We expected co-presence and social presence to be particularly enhanced when observers are co-located with a target person within the same spatial-temporal continuum. Seeing a target person on a screen within the same immersive environment, on the other hand, was expected to induce lower degrees of co-presence and social presence while creating similar levels of immersion and spatial presence.

Expressive body movement and emotion perception

Inferring other’s emotions from their non-verbal behaviors is crucial for navigating our social environment and adjusting

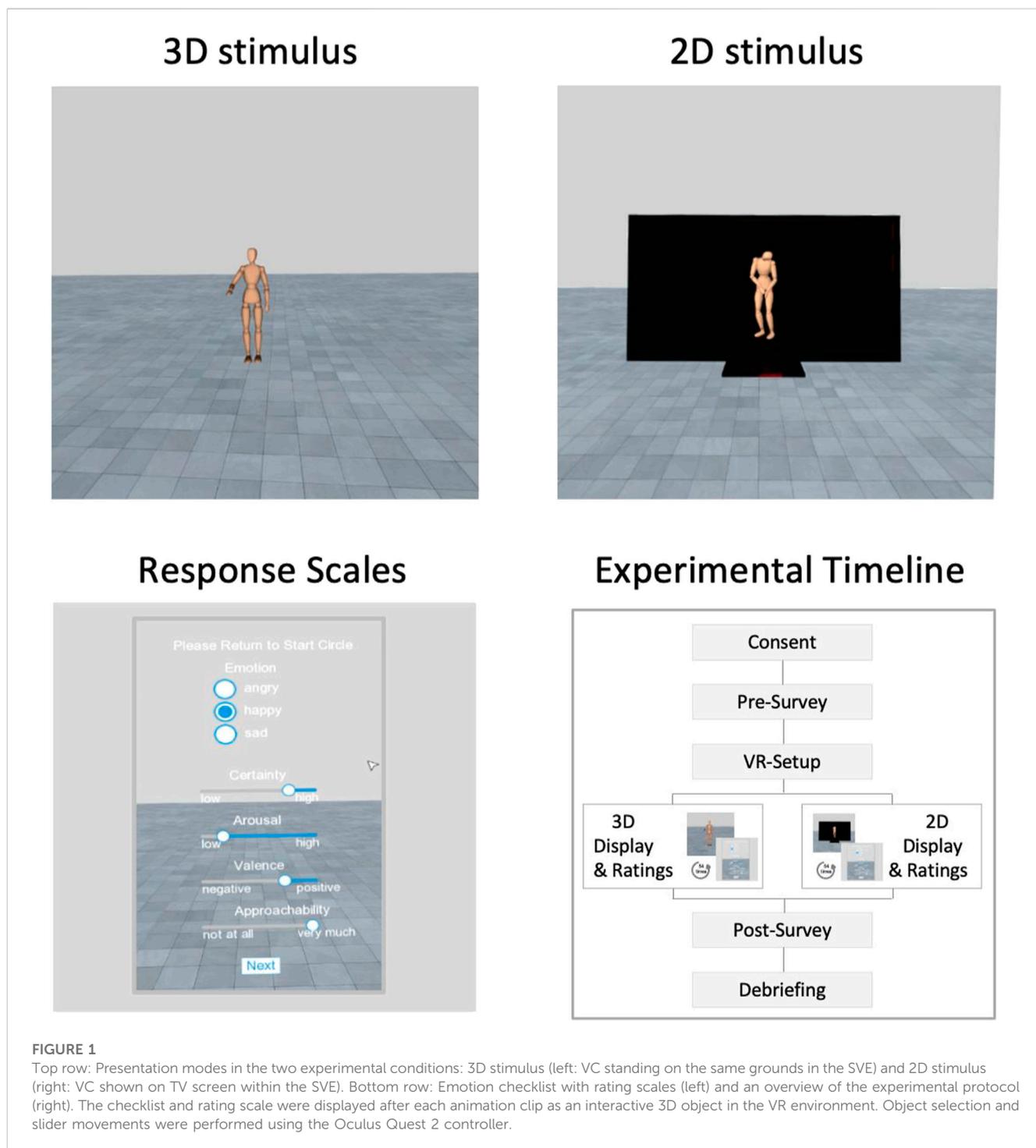
our behaviors in social interactions (Matsumoto et al., 2012). Recent research has provided evidence that expressive body movement (EBM) is a most relevant information source, and that the body alone can provide sufficient cues to differentiate basic emotions (Atkinson et al., 2004; Aviezer et al., 2012; Crane and Gross, 2013; Normoyle et al., 2013; de Gelder et al., 2015). However, several studies have also shown that the accuracy of emotion inferences from EBM varies across emotions, measures and contextual factors (Bente et al., 2001; De Gelder and Van den Stock, 2011; Visch et al., 2014; Martinez et al., 2016). To allow better control of contextual factors as well as to avoid confounds in the stimulus materials, in particular between physical appearance of the targets and their expressive behavior, recent research has increasingly used animated characters to produce lively, yet highly controlled social stimuli (see Bente, 2019).

As recently shown, the recognition of emotional states from VCs’ movements is not a trivial task and emotion recognition often fails. For instance, Reynolds et al. (2019) found significant inter-observer correlations in emotion perception, but the analysis of unbiased hit rates (Wagner, 1993), i.e., match between emotions shown and emotions perceived, revealed difficulties in differentiating between both emotions. In fact, the study was conducted as an online experiment showing the expressive motion sequences as replays on a computer screen. In light of the ecological approach, however, low recognition performance might well be caused by the lack of co-presence and an attenuation of affordances in terms of action possibilities and demands. Against this background, subtle variations of EBM can be considered an ideal candidate to put the role of co-location for emotion perception to the test.

The current study

In sum, the ecological perspective on social perception promotes a basic assumption regarding emotion perception within SVEs: Affordances of expressive behavior are more intensely sensed when the observer and target are co-located in the same space and the behavior of the target can physically impact the observer. As far as IVET is able to create the illusion of proximity and physical impact (for instance, the simple anticipation of a collision) this should also apply to virtual encounters. Sharing the space with a VC, or “standing on the same grounds,” can then be expected to enhance the impact of a VC’s expressive non-verbal behaviors on the observer. More specifically, compared to 2D replays of behavior recordings on a screen, we expected expressive behaviors shown within a co-presence enabling IVET, to disambiguate subtle emotional expressions (see Reynolds, et al., 2019), and thus improve emotion recognition measured as the hit rates and observer agreement as well as the accuracy in discriminating between different emotional qualities. Moreover, in terms of basic behavioral affordances (approach and avoidance), we expected the perceived approachability of the VC to be more distinct for the three emotions (angry, happy, sad) when observed in shared space instead of “on screen.”

The current study puts this assumption to the test, comparing the perceptions of emotionally laden motor activities displayed by a VC either as “offline” replay on a 2D screen, or as “online” animations in a shared space. Importantly, this study did not intend to test the difference between display devices such as screen vs VR headset. Moreover, we aimed to



investigate the difference between expressive behavior displayed by a VC placed in the same spatial-temporal continuum as the observer in contrast to a mediated replay of a VC showing this behavior on a screen. To avoid confounds between this independent variable and potential influences of the media technology (see for instance Hepperle et al., 2020), we placed the replay screen inside the same SVE as the *in situ* appearing VC.

To avoid ceiling effects in emotion recognition due to the use of explicit expressive gestures, e.g., a fist clench (see for instance Aviezer

et al., 2012; Reynolds et al., 2019), we used stimulus material from recently published animation database ACASS (Lammers et al., 2019). The database consists of motion capture recordings of actors performing manual tasks (such as painting, mopping, sanding) in different moods (angry, happy, sad). Having the actors performing a motor task blocks out the use of explicit gestures and reduces expressive component of behavior to implicit dynamics.

TABLE 1 Post-experimental measures.

Measures (adapted from)	# of items: example	scale levels
Manipulation Check (ad-hoc item)	single item: "The avatars I saw were standing on the same floor as me in the virtual environment"	1-I agree 2-neither/nor 3-I do not agree
VR Sickness (Kim et al., 2018)	9 symptom items: e.g., "general discomfort," "eye-strain," "dizziness"	1-none 2-slight 3-moderate 4-severe
Controller Naturalness (McGloin et al., 2011)	12 items: "The virtual environment controls seemed natural"	
Realism (Poeschl and Doering, 2013)	6 items: "I felt that the avatar displayed the actions of a real person"	
Immersion (Witmer and Singer, 1998)	6 items: "I forgot about my real environment during the experiment"	1-strongly disagree 2-disagree 3-neither/nor
Co-Presence (Bente et al., 2008, + Slater et al., 2000)	6 items: "It felt as though the avatar was with me in the room"	4-agree 5-strongly agree
Spatial Presence (Hartmann et al., 2016)	6 items: "It felt as though I was physically present in the virtual environment."	
Social Presence (Bente et al., 2008, + Lombard et al., 2011)	6 items: "I was aware of the moods the avatars were in"	

Methods and materials

Participants

The study sample comprised 94 students (*mean age* = 19.8, *sd* = 1.77, 61.3% female) from a Midwestern University. All participants provided written consent to the protocol, which was approved by the local review board, and they received course credit or a 10 US\$ incentive.

Stimulus material

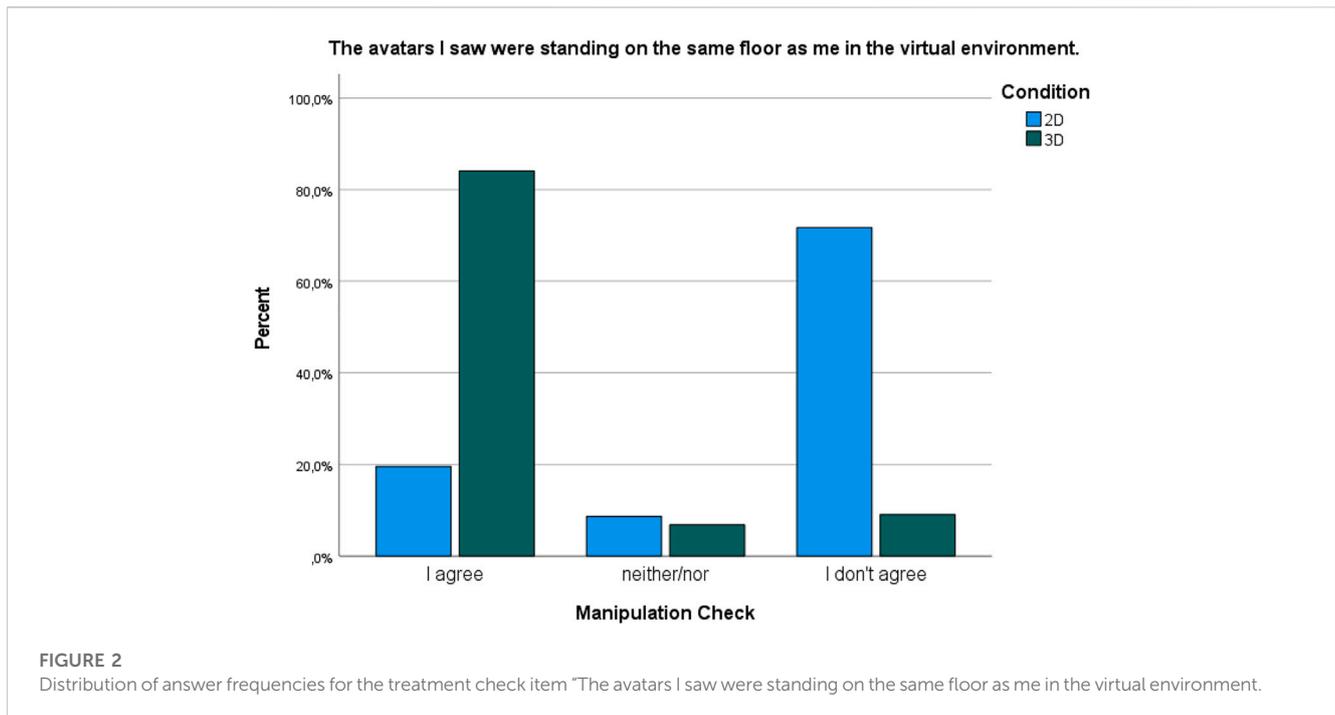
Animation stimuli were taken from a recently published database ACASS (Annotated Character Animation Stimulus Set; Lammers et al., 2019). ACASS consists of motion capture datasets of human actors performing six different everyday activities (sweeping, wiping a table, painting with a roller, painting with a brush, mopping, sanding a table) in three different moods (angry, happy, sad). The animations are available as rendered MP4 videos (1,920 × 1,080 resolution) usable for screen presentation and as animation data files in FBX format, which allow real time 3D rendering on a VR display. FBX files contain the mocap data collected with an optical motion capture system (Optitrack™, NaturalPoint, Inc., Oregon, United States) mapped onto a simple virtual character (wooden mannequin). Stimulus features and production details can be found in Lammers et al. (2019). Of note, the VC animations didn't show facial expressions (see Figure 1).

From the ACASS database, we selected a total of 54 experimental stimulus clips that comprised three emotions (angry, happy, sad), and three of the overall six activities (mopping, paint-brushing, sanding). For exercise purposes two further clips were randomly selected from activity categories that were not included in experimental stimuli (sweeping, paint-roller). The stimulus set also contained three repetitions for each of 18 demographically different actors as well as variations of the start camera angle (slight from the right or left). Stimuli were selected from the ACASS database using the following procedure: 1) nine clips from each emotion x activity category were randomly selected, 2) the researcher inspected these clips to make sure no more than three out of 54 total clips were selected from the same actor and any fourth clip selected from the same actor was replaced by the next clip from the same emotion x activity category in the randomized list. 3) Half of the final selections for each emotion x activity category were then randomly assigned to either left or right camera angle presentation.

From the selected set of ACASS clips, we used the MP4 versions for the 2D condition and the FBX versions for the 3D condition. 2D, as well as the 3D stimuli, were both presented within the same SVE that was displayed via the Oculus/Meta Quest 2 VR headset. A python-based VR programming environment (Vizard, 7.0; Worldviz, 2021) was used to create the SVE and either run the real-time animation of the FBX-data or the replay of the MP4 clips (H.264 encoded) on the virtual TV screen.

Measures and procedures

The timeline of the experiment is shown in Figure 1. Upon arrival, participants were introduced to the experimental setup,



provided informed consent, and filled out a pre-experimental Qualtrics survey on a PC asking for general demographic data.¹ After completion, they were helped to put on the VR headset. The presentation program was started, and the participants of both groups entered the SVE environment that allowed for locomotion in about a 2-m radius within the scene. Participants in the 3D condition saw the VC acting in the same shared space, whereas participants in the 2D condition saw the MP4-version of the same VC animation on a virtual TV screen that was placed within the SVE at the same initial distance as the in-situ VC (see Figure 1). Thus, the technical presentation features were held constant, including the wearing of a VR headset and the possibility to move in the VR environment. After each 10-s stimulus clip, participants were asked to categorize the emotion observed (angry, happy, or sad) and to quantify subjective judgment certainty (low-high), perceived arousal (low-high), and valence (positive-negative), as well as approachability of the VC (not at all - very much). Radio buttons for the forced choice categorization and sliders for the ratings were presented within the SVE so that

participants did not have to take off the VR headset after each stimulus presentation (see Figure 1 lower left). Participants used the right Quest 2 controller to click the radio buttons and to adjust the sliders. After completion of stimulus presentation, participants took off the headset and completed a post-experimental Qualtrics survey regarding their VR experience.

Table 1 gives an overview of the post-hoc survey scales, including the respective references, number of rating items, their polarity and range, as well as example items. In the items referring to the perception of VCs we used the term avatar as this term is more familiar to most people. The survey included a one-item manipulation check regarding the observer's perspective as perceived in the 2D and 3D condition, asking the agreement to the statement 'I had the impression to stand on the same floor as the avatar', (1= I agree, 2 = I neither agree nor disagree, not agree, 3= I do not agree). We further included technology oriented questions regarding immersion, realism and controller naturalness as well as a scale for possible negative effects of the immersive technology in terms of VR-sickness. More specifically, we asked about the participant's presence experience differentiating between spatial presence, co-presence and social presence. Consistent with previous usage, rating polarities were held constant across all presence scales asking for levels of agreement ranging from "1" = "I strongly disagree" to "5" = "I strongly agree." The wording of the scales was adapted to specifically target the perception of a VC's expressive behaviors. Also, for the immersion and presence scales, subsets of items (up to six) were selected from the original scales to keep the survey as short and concise as possible. Scales and items can be found in the **Supplementary Material**. After completion of the post-experimental survey, participants were debriefed and received the incentive.

¹ Additional data were collected in this study that are not reported in this paper. These include survey questions about participant's general IVET interest, familiarity, and usage, the Immersive Tendencies Questionnaire (Witmer and Singer, 1998) as well as the Positive Affect Negative Affect Scale (PANAS, Watson et al., 1988) that was applied before and after the stimulus presentation phase. Furthermore, physiological measures of attention and arousal (EEG, cardiovascular activity) were captured during the whole stimulus presentation sequence using a Muse 2 headset device, Interaxon Inc. (SCR_014418) worn under the VR headset. Furthermore, 6-degrees-of-freedom head motion data was recorded from the Quest 2 headset sensors. These data will be reported in a separate publication focusing on the emotional effects of expressive body movement on the observers.

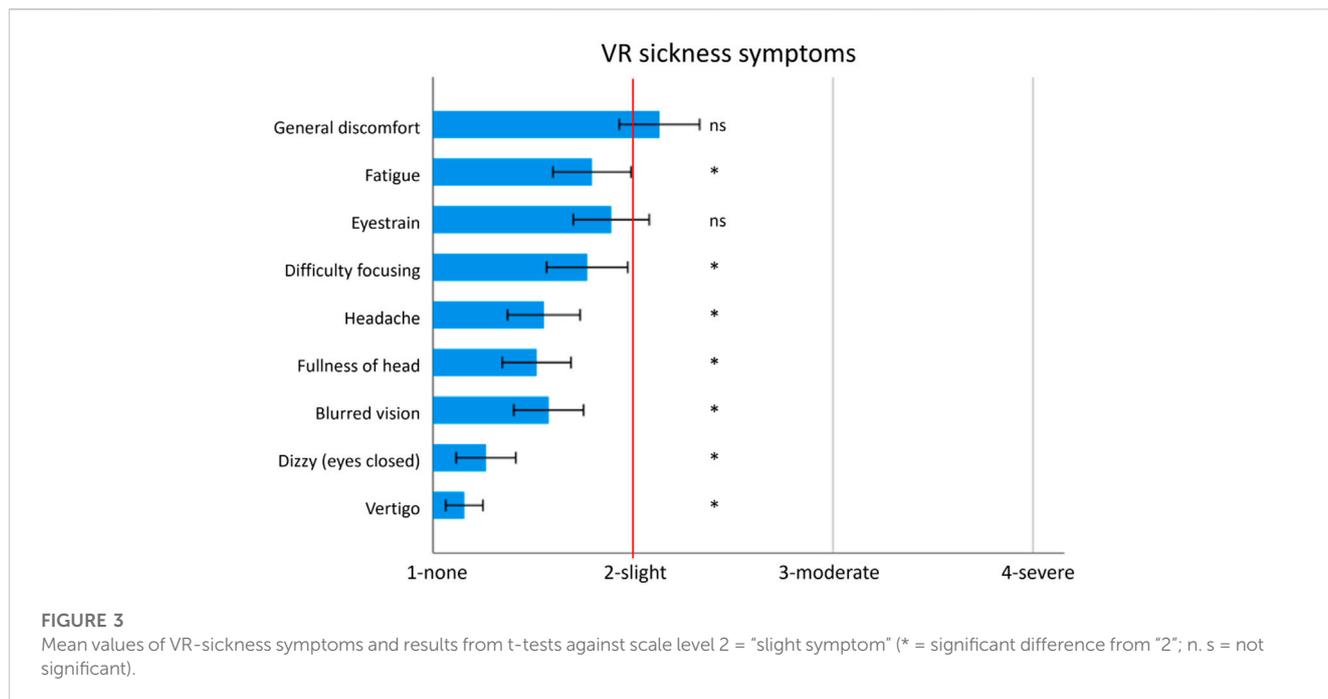


TABLE 2 t-test comparisons for perceived realism, immersion, spatial presence, co-presence, and social presence.

Measures	2D Mean (SD)	3D Mean (SD)	t (df)	p	Cohen's d
Perceived Realism	3.06 (.68)	3.30 (.53)	-1.01 (87)	.314	-.215
Immersion	2.91 (.76)	3.05 (.61)	-.99 (88)	.325	-.209
Spatial Presence	3.50 (.74)	3.65 (.67)	-1.96 (88)	.054	-.41
Co-Presence	2.76 (.88)	3.21 (.66)	-2.77 (89)	.007	-.58
Social Presence	2.99 (.67)	3.23 (.44)	-1.98 (88)	.052	-.416

Results

Treatment check

A treatment check regarding the perceived co-location of observer and VC was performed asking participants to indicate their agreement to the statement "I had the impression to stand on the same floor as the avatar" (1= I agree, 2= I neither agree or, not agree, 3= I do not agree). A Chi-square test regarding the answer distribution of the treatment check item revealed a highly significant difference between the two experimental conditions ($\chi^2 (2, 90) = 39.89, p < .001$). Figure 2 illustrates the result, showing that the VC presentation in the co-location condition was perceived as standing on the same floor while the VC displayed on the screen was not.

General effects of VR usage

To examine whether the use of VR technology impacted participant's comfort level, we analyzed the responses to the VR sickness questionnaire (VRSQ). A one-sample t-test against scale level "2" (slight symptom) was conducted for all 9 items. As shown in Figure 3, only the items "general

discomfort" and "eyestrain" didn't deviate significantly from scale level 2, indicating that mild forms of these symptoms were experienced by the participants. All other items deviated significantly from scale level 2 in the direction of "none," indicating that these symptoms were absent or only minimally experienced by the participants.

Statistical tests for the immersion, realism and controller naturalness scales were based on the averages across items. A one-sample t-test comparison for controller naturalness (mean score over all participants = 3.14) revealed a significant deviation from the scale's mean of 3.0 ($t_{(79)} = 2.47, p < .016, d = .276$, see Table 2). Except for one item that was inverted before calculating averages, all items were positively formulated. Smaller numbers indicated that participants disagreed, i.e., participants feel slightly positive about the controller handling and navigation. No significant difference in controller naturalness was found between the two groups ($t_{(78)} = -.50, p = .62$).

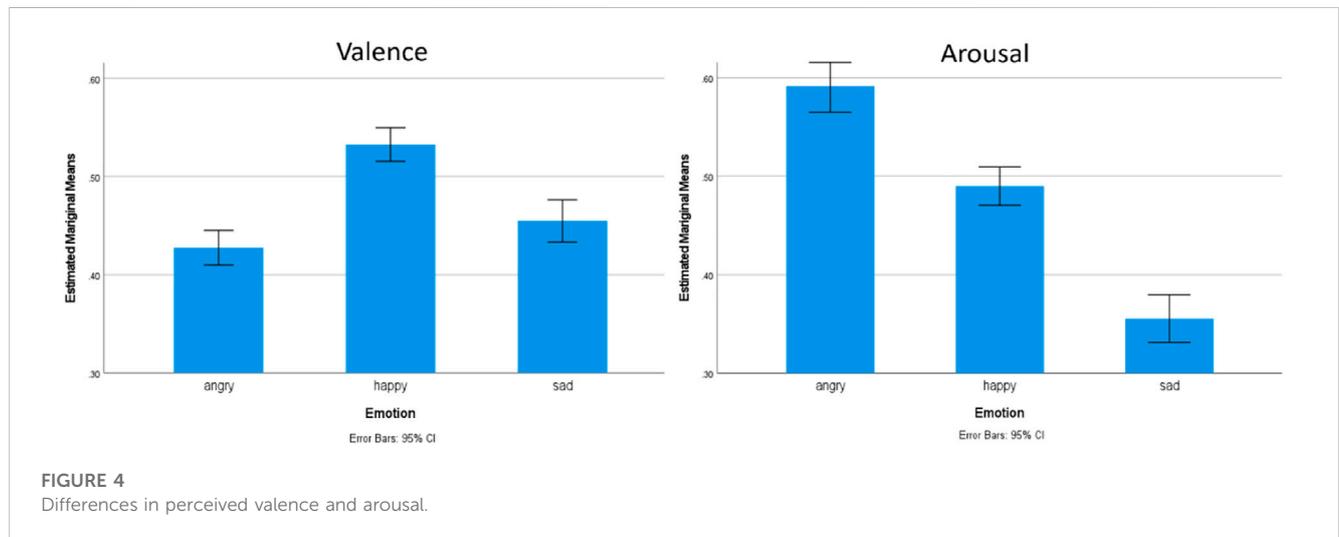
Also, no significant differences between both groups were found for perceived realism ($t_{(87)} = -1.01, p = .314$) and for immersion ($t_{(87)} = -.99, p = .325$). While the overall mean for immersion did not significantly deviate from the scale mean "3" (neither agree nor disagree) ($t_{(80)} = .795, p = .429$), perceived realism showed a significant difference in the positive direction, yet with a small effect size ($t_{(80)} = 2.138, p = .035, d = 0.233$).

TABLE 3 Mean scores for unbiased hit rates (Hu), judgment certainty, perceived valence, arousal of the EBM, and approachability of the VC.

Measure	Angry		Happy		Sad	
	2D Mean (SD)	3D Mean (SD)	2D Mean (SD)	3D Mean (SD)	2D Mean (SD)	3D Mean (SD)
<i>Hu</i>	.44 (.16)	.46 (.13)	.21 (.10)	0.27 (.12)	.37 (.14)	.43 (.13)
<i>Certainty</i>	.60 (.11)	.58 (.11)	.55 (0.10)	0.56 (.13)	.58 (.10)	.59 (.12)
<i>Valence</i>	.42 (.09)	.44 (.07)	.52 (0.08)	0.54 (.08)	.46 (.10)	.45 (.10)
<i>Arousal</i>	.61 (.11)	.57 (.13)	.49 (0.08)	0.49 (.10)	.34 (.11)	.37 (.12)
<i>Approachability</i>	.43 (.10)	.45 (.10)	.58 (0.11)	0.59 (.11)	.55 (.12)	.53 (.13)

TABLE 4 Results from repeated measurement ANOVA with emotion as within-subject factor and stimulus modality as between-subject factor.

Measure	F _{Emotion}	p	partial η^2	F _{Stimulus Modality}	p	partial η^2	F _{Emotion x Stimulus Modality}	p	partial η^2
<i>Hu</i>	95.97	<.001	.539	3.69	.058	.043	1.06	.350	.013
<i>Certainty</i>	6.63	.002	.074	.01	.918	.000	2.59	.078	.03
<i>Valence</i>	60.48	<.001	.422	.61	.437	.007	2.0	.138	.024
<i>Arousal</i>	134.9	<.001	.619	.12	.726	.001	2.18	.135	.026
<i>Approachability</i>	81.63	<.001	.496	.06	.816	.001	2.43	.091	.028



Differences in perceived presence

Cronbach’s α was calculated for the three ad-hoc presence scales revealing satisfactory reliability (spatial presence: $\alpha = .76$; co-presence = .80; social presence = .68). Based on the means across the six items per scale, *t*-test comparisons were conducted to identify differences between the experimental groups. Table 2 also shows the results for the three scales. As visible, *co-presence*, i.e., the feeling of being in the same space as the VC, showed a significant difference in

the expected direction, whereas *spatial presence* and *social presence* only showed marginal effects in the expected direction, i.e., stronger effects in the 3D condition.

Differences in emotion perception

Next, we examined differences in emotion perception between the experimental groups. To correct for response biases towards one

of the three emotions, we used *unbiased hit rates* (*Hu*, see Wagner, 1993) instead of absolute hit rates. The *Hu*-scores were calculated for each participant across the 54 stimuli (18 within each emotion category) and ratings of certainty, valence, arousal, and approachability were averaged across the 18 stimuli within each emotion category for each participant. Table 3 shows the averages for each condition.

To statistically test for differences between the conditions, we conducted repeated measurement ANOVAs for the dependent variables *unbiased hit rate* (*Hu*, arcsin transformed for ANOVA and *t*-tests) and each of the rating dimensions. *Display condition* (2D versus 3D) was entered as a between-subjects factor and *displayed emotion* (angry, happy, sad) as within-subjects factor. Table 4 shows the results of these analyses. While the within-subject factor *emotion* showed significant main effects across all five parameters, the between-subject factor *stimulus modality* (2D vs. 3D) only revealed a marginally significant difference for the *unbiased hit rate* (*Hu*). No significant interaction effects were found.

Separate post-hoc *t*-test comparisons of the parameter *Hu* for the three emotions then revealed a significant difference between the *stimulus modalities* for the emotion *happy* ($t(82) = 2.29, p = .025, d = .500$). In the 3D condition, the intended happy expressions were more likely to be recognized as happy by the observers than in the 2D condition (means see table 4). Importantly though, *Hu*-scores in both conditions remained below chance level. A marginally significant effect of the experimental condition was also found for the emotion *sad* ($t(82) = 1.865, p = .066, d = .407$). In the 3D condition, the sad expressions were more likely to be recognized as such (means see table 4). Overall the emotion *sad* had the best hit rates (above chance level) and also slightly benefitted from the 3D display.

Additional analyses for emotion perception

Consistent with earlier findings, the overall unbiased hit rates (*Hu*-scores) in this study were low (see Reynolds, et al., 2019), and only the *Hu*-scores for anger significantly exceeded the chance level. Because of this lack of correspondence between *emotions shown* (actor instruction) and the *emotions perceived*, we further explored if the observers were able at all to differentiate between the three emotional qualities and how well they agreed in their judgments even if these were not matching the displays. Of note, in these analyses, we collapsed data over both display modalities. Consistent with earlier research, we found significant differences between the three emotions in perceived *valence* ($F(1.958, 164.44) = 60.48, p < 0.001, \eta_p^2 = .422$) and *arousal* ($F(1.334, 112.05) = 134.9, p < 0.001, \eta_p^2 = .619$). Figure 4 illustrates this finding. As can be seen, *valence* was lowest for the *angry* and highest for *happy* stimuli with *sad* in between. *Arousal* was lowest for the *sad* and highest for *angry* stimuli with *happy* in between. Thus, while participants failed at category assignment, their ratings of valence and arousal reveal significant and expected differences in the perception of these three emotions.

The low hit rates, as reported above, suggest that raters weren't very successful in identifying the VCs expressed emotion. This raises the question of whether participants were just performing at chance

level or whether there was a systematic error, such as when participants would consistently mistake, e.g., happiness for anger. To examine this question, we assessed the degree to which the observers agreed in their judgment. Specifically, we tested whether participants consistently categorized the emotions (using kappa nominal agreement statistic; Fleiss, 1971) as well as whether they converged in their continuous ratings of certainty, valence, arousal, and approachability (using intra-class coefficients; Shrout and Fleiss, 1979). For the emotion identification task, we find an overall agreement $\kappa = .357$, suggesting moderate, but significant agreement among raters. Similarly, participant's continuous evaluations of the 54 stimuli also exhibited high levels of agreement across observers (Certainty: $ICC_{3,k} = .9$; Valence: $ICC_{3,k} = .9$; Arousal: $ICC_{3,k} = .97$; Approachability: $ICC_{3,k} = .89$). Not only did the observers show high levels of agreement, but they also individually felt certain about their judgments. One sample *t*-tests for *certainty* ratings revealed for all three emotions significant positive deviations from the scale mean (*angry*: $t_{(84)} = 7.645, p < .001, d = .829$; *happy*: $t_{(84)} = 4.419, p < .001, d = .479$; *sad*: $t_{(84)} = 6.903, p < .001, d = .749$).

Discussion

Inferring others emotions from their non-verbal behavior is a crucial element in social interactions (see Matsumoto et al., 2012). It is therefore not surprising that VCs, including avatars and Embodied Conversational Agents (ECAs), are supposed to fundamentally change the way we communicate via VR media of the future, and also how we communicate with intelligent machines. While the possibilities to capture and transmit human motion in full bandwidth are still very limited with today's commercial devices, research has long discovered the unique possibilities that VR technologies offer for the study of emotion perception in the laboratory (see Bente, 2019). Full body motion capture systems and professional character animation tools allow for meticulous tracking of human motion and its experimental variation and replay *via* realistic VCs. However, most person perception studies using VCs have presented the character animations on a screen, while immersive VR displays have rarely been used in this research domain. Surprisingly, the influence of display modalities on emotion perception has not been tested nor discussed so far. We here addressed this question expecting the answer to be essential for evaluating previous results and understanding contextual effects on emotion perception in more general terms.

The current study was based on an ecological perspective placing social affordances, i.e., needs and options for (re-)action, at the core of emotion perception. We specifically reasoned that sharing the same space with an animated VC should amplify social affordances as compared to the offline replay of VC animations on a screen. We specifically expected that co-location of observer and target in an SVE would lead to a more intense feeling of co-presence and social presence and consequently alter observers' emotion recognition performance, their felt judgment certainty, and the perceived intensities of valence and arousal. In a between-subject design, we compared the perception of 54 10-s animations of VCs performing three daily activities (painting, mopping, sanding) in three

emotional states (angry, happy, sad). While one group saw the life-size VC performing the action in-situ (3D condition), the second group saw the VC animations on a virtual TV screen placed within the same SVE. We thus held all setting factors (SVE, VR headset, controllers) constant across conditions, allowing us to attribute differences solely to the experimental variable *co-location*, i.e., in-situ vs on-screen, except for those differences in physical stimulus characteristics (such as color) that are inherently necessary to convey the 3D information of the VC animations as opposed to their 2D counterparts and thus non-separable (see below). Accordingly, we expected no differences between the experimental groups in the level of immersion, realism, and spatial presence they felt in the SVE.

Main results: Experimental effects

Regarding the future use of similar setups, it was important to show that the experimental induction worked and that wearing the VR technology and the immersive experience were not perceived as inconvenient or causing symptoms of VR-sickness. Results provide strong evidence that the experimental manipulation of presentation modalities was successful: Participants in the fully embodied VC condition (3D) consistently reported that they had the impression of standing on the same ground with the VC, while participants in the 2D screen condition didn't. Importantly, participants did not experience any VR sickness symptoms worth noting during the approximately 20-min session, at least not beyond minimal symptoms of general discomfort and eye strain. No stress, fatigue or dizziness was reported in any of the conditions.

Analyzing the general effects of VR, we found that participants were only slightly positive about controller naturalness. The reserved evaluation might be explained by the fact that the controller was not used for navigation, but as a mouse surrogate to select radio buttons and adjust sliders for the emotion ratings. As expected, both experimental conditions also did not differ with regard to perceived realism and immersion. However, immersion didn't deviate significantly from the neutral scale mean (neither agree or disagree), and perceived realism only showed a weak effect in the expected direction. The weakly positive response to the realism scale might be due to the use of an abstract wooden manikin figure as VC. Overall, these data suggest that there is room for improvement regarding user engagement within the SVE. One aspect might be the layout of the space that was just displayed as an open plane, i.e., without walls or other orientation points. More importantly, the lack of realism might have been caused by the VC's physical appearance, as a wooden manikin.

We further expected the experimental conditions to differentially impact the participants' experiences of *spatial presence*, *co-presence*, and *social presence*. Specifically, we expected the 3D condition to induce significantly higher levels of *co-presence* and *social presence*. We expected no differences in *spatial presence*, though, as both groups entered the same SVE with the same locomotion possibilities. Consistent with our expectations, we found a significant difference for *co-presence*. Results for *social presence* were marginally significant, as were the results for *spatial presence*. The *co-location* of observer and VC thus was clearly reflected in the participant's *co-presence* ratings and also showed a tendency to positively influence the feeling of

psychological connectedness (*social presence*). Surprisingly, and in contrast to our assumptions, *spatial presence*, i.e., the feeling of being transported to the virtual place, also showed a consistent tendency. The reason might be that the presence of a social entity, i.e., the population of an otherwise empty space, defines that space as a social place and possibly induces social affordances, in terms of approach and avoidance, i.e., locomotion towards or away from the VC. This question should receive further attention in future studies. Moreover, as briefly indicated above, the presentation of the same item as 3D as opposed to 2D is inherently associated with a few lower-level sensory differences that convey the 3D information. Although we consider it unlikely that these factors could explain the current results, the possibility cannot be entirely discarded. As such, additional studies (e.g., *via* black-white presentations, or different lightning conditions) could also be carried out.

Clearly though, *co-presence* was the only dimension that revealed a significant experimental effect. In our view, the specificity of the experimental manipulation effects is quite remarkable as all setting features were held constant across the experimental conditions (same VR technology, same stimuli, same room inside VR) except the presentation of the VC motion either as in-situ animation or as screen-based replay. There is consensus that the boundaries between *spatial presence*, *co-presence*, *social presence*, are blurred and difficult to distinguish *via* respective rating items, such as "I felt as though I could actually reach out and touch the avatar in the virtual environment," "It felt as though the avatar was with me in the room," "I was aware of the avatars' moods" (see Bulu, 2012; Cummings and Bailenson, 2016). All the more, it is important to point out that the ad-hoc scales used in this study were sensitive to our experimental variation and offer themselves as short presence survey scales for future studies.

In contrast to the presence data, no systematic effects of the experimental variation could be found in either one of the four rating dimensions (arousal, valence, certainty, and approachability). A significant effect between 2D and 3D conditions was found in the hit rates (accordance between emotion shown and emotion named) for the emotion "happy." Here the unbiased hit rates (*Hu*) were significantly higher in the 3D condition. A marginal effect on the hit rates in the expected direction was also found for the emotion "sad." As there was only a small effect size for both emotions and no other evidence pointed in the same direction we refrain from further interpretation. Importantly, for the emotion "happy," the *Hu*-scores in the 3D condition still didn't surpass chance level, which suggests that an overall pronounced judgment error might have been slightly alleviated in the 3D condition, but recognition still didn't really benefit in terms of "correct" judgments. In contrast, hit rates for sadness were also low but above chance level in both conditions, suggesting that sadness in general was easier to detect.

One possible explanation for the overall low hit rates and the absence of differences in emotion perception between the experimental conditions lies in the nature of the stimulus materials, which may not have been optimal for expressing the phenomenon of interest. Affordances, according to Gibson, are what a stimulus "offers" to a person in terms of action, i.e., how it can be used. In this sense, "*perception is for action*" (Zebrowitz, 2002). Applied to emotion perception, this view differs from a more common view in emotion psychology as well as affective

computing, which tends to assume that a target's internal emotional state is the ground truth and that emotions are "detected" or "read out" (see for instance Buck, 1985). Ecological or action-oriented models of emotion perception paint a far more dynamic picture of the situation—emphasizing the possibilities to re-act to an expressive behavior, for instance in basic terms of approach and avoidance. In the current study, however, participants observed VCs performing manual activities at a more or less constant distance. There was no task or a perceivable need to move toward the VC or prepare for interaction. This raises the question of whether effects would be stronger if the VCs would approach the observers or observers would be asked to actively approach the VC or even interact with the virtual other.

Somewhat related arguments have been made in the realm of physiological and clinical literature, that has demonstrated how the "looming" nature of emotional, in particular aversive and fear inducing stimuli, and the anticipation of collision, modulates defensive responses in the observer (see for instance Huijsmans et al., 2022). These responses could be conceptualized in terms of affordances, and behavioral as well as neurophysiological measures could help to further understand the specific perception-action links in virtual social encounters. In fact, although Zebrowitz (2002) claimed the importance of IVETs in studying social affordances, up to now only few efforts have been made to exploit the potential of immersive technologies to this end systematically. A reason might be that only recently the VR technology has become affordable and easier to use. While all VR headsets already include sensors to track users' locomotion, newer systems such as the *HP Reverb Omnicast* and *Meta Quest Pro* also provide sensors to measure gaze, cardio-activity and lower face movements. Furthermore, available software tools, such as the Vizard platform used here, offer flexible possibilities to tailor SVEs and VCs to specific research needs. Overall, the VR instrumentation now at hand allows studying social affordances across a number of cognitive, emotional and behavioral variables, and even in rich virtual environments that may include interactive VCs like avatars and ECAs.

Additional analyses: Emotion perception

Additional analyses of the emotion perception data revealed further insights into the processing of VC's expressive body movements (EBM) that deserve special attention. Recognition rates (unbiased hit rates, *Hu*) were only slightly above chance level for *angry* and *sad* and even below chance level for *happy*. At the first glance these results suggest that the stimuli were ambiguous and not at very informative or the participants were not motivated to invest much effort in the recognition task. In both cases the observers category assignments could be expected to be randomly distributed across the three categories, and the ratings of arousal, valence and approachability should either regress to the scale means (indifferent) or randomly vary as well while certainty rating should be low. In particular, one couldn't at all expect significant agreement between the observers. In fact, the opposite was the case: Observers reported significant levels of certainty as tested toward the scale mean. They showed significant, yet moderate agreement in category assignment and reached impressive levels of

agreement in valence, arousal and approachability ratings. Furthermore, while category assignment was mostly erroneous, average ratings of arousal and valence significantly, and consistently with earlier research, differentiated between the emotions and revealed the expected level differences, i.e., "*happy*" being the most positive, followed by "*sad*" and then by "*angry*," and "*angry*" being the most arousing being followed by *happy* and then by "*sad*." So, the implicit qualities of the *emotions shown* were consistently perceived, but the category assignment failed.

The stimulus subsets representing the three emotions were evidently heterogeneous, as, putatively there is also a considerable variance in EBM in everyday life. Interestingly, the subtle information conveyed *via* movement dynamics still induced consistent ratings across observers. This leads to the question on which implicit motion qualities observer judgments are drawing upon. As the underlying motion capture data driving the VC's behavior is available to us, further analyses are intended to parametrize EBM, for instance in terms of velocity, acceleration, and expansivity, and relate this data to the observer judgments, focusing the analysis on the *emotions perceived* instead of the *emotion shown*. (See Reynolds et al., 2019).

Lastly, future studies could use the inherent possibilities of VR systems to track observer motion by looking at the dynamics of approach and avoidance behaviors relative to the target and to virtual objects or other social entities in the SVE. These kinds of experimental arrangements would most perfectly match the requirements for studying social affordances as laid out in earlier work (see Zebrowitz, 2002).

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 Michigan State University Internal Review Board. The patients/participants provided their written informed consent to participate in this study.

Author contributions

GB is the principle investigator and project lead from inception to completion. RS was principle in making the study operational, data analysis, and writing of the paper. NJ aided in making the study operational, data collection, data presentation, and writing of the paper. AS aided in study design, operationalization, data collection, analysis, project write up.

Funding

This work was supported by the National Science Foundation (grant number 1907807).

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

References

- 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 (6), 717–746. doi:10.1068/p5096
- Aviezer, H., Trope, Y., and Todorov, A. (2012). Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science* 338 (6111), 1225–1229. doi:10.1126/science.1224313
- Axelsson, A. S., Abelin, Å., Haldal, I., Schroeder, R., and Wideström, J. (2001). Cubes in the cube: A comparison of a puzzle-solving task in a virtual and a real environment. *CyberPsychology Behav.* 4 (2), 279–286. doi:10.1089/109493101300117956
- Bailenson, J. (2018). *Experience on demand: What virtual reality is, how it works, and what it can do*. New York, United States: W. W. Norton & Company.
- Bailenson, J. N., Blascovich, J., Beall, A. C., and Loomis, J. M. (2003). Interpersonal distance in immersive virtual environments. *Personality Soc. Psychol. Bull.* 29 (7), 819–833. doi:10.1177/0146167203029007002
- Basori, A. H., and Ali, I. R. (2013). Emotion expression of avatar through eye behaviors, lip synchronization and MPEG4 in virtual reality based on Xface toolkit: Present and future. *Procedia - Soc. Behav. Sci.* 97, 700–706. doi:10.1016/j.sbspro.2013.10.290
- Bente, G., Krämer, N. C., Eschenburg, N., and Eschenburg, F. (2008). Avatar-mediated networking: Increasing social presence and interpersonal trust in net-based collaborations. *J. Hum. Res. Rehabilitation* 34 (2), 287–318. doi:10.1111/j.1468-2958.2008.00322.x
- Bente, G., Krämer, N. C., Petersen, A., and de Ruiter, J. P. (2001). Computer animated movement and person perception: Methodological advances in nonverbal behavior research. *J. Nonverbal Behav.* 25 (3), 151–166. doi:10.1023/A:1010690525717
- Bente, G., Krämer, N. C., Trogemann, G., Piesk, J., and Fischer, O. (2001). “Conversing with electronic devices. An integrated approach towards the generation and evaluation of nonverbal behavior in face-to-face like interface agents,” in *Intelligent interactive assistance and mobile multimedia computing. Proceedings of the workshop IMC 2000*. Editors A. Heuer and T. Kirste (Rostock: Neuer Hochschulschriftenverlag), 67–76.
- Bente, G. (2019). “New tools – new insights: Using emergent technologies in nonverbal communication research,” in *Reflections on interpersonal communication*. Editors S. W. Wilson and S. W. Smith (Solana Beach, CA: Cognella Academic Publishing), 161–188.
- Bente, G., Rüggenberg, S., Krämer, N. C., and Eschenburg, F. (2008). Avatar-mediated networking: Increasing social presence and interpersonal trust in net-based collaborations. *Hum. Commun. Res.* 34 (2), 287–318. doi:10.1111/j.1468-2958.2008.00322.x
- Blascovich, J., and Bailenson, J. (2011). Infinite reality: Avatars, eternal life, new worlds, and the dawn of the virtual revolution. New York: William Morrow & Co.
- Buck, R. (1985). Prime theory: An integrated view of motivation and emotion. *Psychol. Rev.* 92 (3), 389–413. doi:10.1037/0033-295x.92.3.389
- Bulu, S. T. (2012). Place presence, social presence, co-presence, and satisfaction in virtual worlds. *Comput. Educ.* 58 (1), 154–161. doi:10.1016/j.compedu.2011.08.024
- Bystrom, K. E., and Barfield, W. (1999). Collaborative task performance for learning using a virtual environment. *Presence* 8 (4), 435–448. doi:10.1162/105474699566323
- Casanueva, J. (2001). *Presence and co-presence in collaborative virtual environments*. M.Sc. Dissertation (South Africa: University of Cape Town).
- Crane, E. A., and Gross, M. M. (2013). Effort-shape characteristics of emotion-related body movement. *J. Nonverbal Behav.* 37 (2), 91–105. doi:10.1007/s10919-013-0144-2
- 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
- De Gelder, B., De Borst, A. W., and Watson, R. (2015). The perception of emotion in body expressions. *Wiley Interdiscip. Rev. Cognitive Sci.* 6 (2), 149–158. doi:10.1002/wcs.1335
- De Gelder, B., and Van den Stock, J. (2011). The bodily expressive action stimulus test (BEAST). Construction and validation of a stimulus basis for measuring perception of whole body expression of emotions. *Front. Psychol.* 2, 181. doi:10.3389/fpsyg.2011.00181
- Dings, R. (2018). Understanding phenomenological differences in how affordances solicit action. An exploration. *Phenom. Cogn. Sci.* 17 (4), 681–691. doi:10.1007/s11097-017-9534-y
- Fabri, M., Moore, D., and Hobbs, D. (2004). Mediating the expression of emotion in educational collaborative virtual environments: An experimental study. *Int. J. Virtual Real.* 7, 66–81. doi:10.1007/s10055-003-0116-7
- Felton, W. M., and Jackson, R. E. (2022). Presence: A review. *Int. J. Human-Computer Interact.* 38, 1–18. doi:10.1080/10447318.2021.1921368
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychol. Bull.* 76 (5), 378–382. doi:10.1037/h0031619
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton Mifflin Company.
- Goffman, E. (1963). *Behavior in public places*. New York: The Free Press.
- Grabarczyk, P., and Pokropski, M. (2016). Perception of affordances and experience of presence in virtual reality. *AVANT J. Philosophical-Interdisciplinary Vanguard*, 25–44. doi:10.26913/70202016.0112.0002
- Hartmann, T., Wirth, W., Schramm, H., Klimmt, C., Vorderer, P., Gysbers, A., et al. (2016). The spatial presence experience scale (SPES). *J. Media Psychol.* 28 (1), 1–15. doi:10.1027/1864-1105/a000137
- Hepperle, D., Ödell, H., and Wölfel, M. (2020). “Differences in the uncanny valley between head-mounted displays and monitors,” in International conference on cyberworlds (CW), Caen, France, 41–48. doi:10.1109/CW49994.2020.00014
- Hepperle, D., Purps, C. F., Deuchler, J., and Wölfel, M. (2022). Aspects of visual avatar appearance: Self-representation, display type, and uncanny valley. *Vis. Comput.* 38, 1227–1244. doi:10.1007/s00371-021-02151-0
- Huijsmans, M. K., de Haan, A. M., Müller, B. C., Dijkerman, H. C., and van Schie, H. T. (2022). Knowledge of collision modulates defensive multisensory responses to looming insects in arachnophobes. *J. Exp. Psychol. Hum. Percept. Perform.* 48 (1), 1–7. doi:10.1037/xhp0000974
- Kim, H. K., Park, J., Choi, Y., and Choe, M. (2018). Virtual reality sickness questionnaire (VRSQ): Motion sickness measurement index in a virtual reality environment. *Appl. Ergon.* 69, 66–73. doi:10.1016/j.apergo.2017.12.016
- Lammers, S., Bente, G., Tepest, R., Roth, D., and Vogeley, K. (2019). Introducing ACASS: An Annotated Character Animation Stimulus Set for controlled (e)motion perception studies. *Front. Robotics AI* 6, 94. doi:10.3389/frobt.2019.00094
- Lee, K. M. (2004). Presence, explicated. *Communication Theory: CT: A Journal of the International Communication Association* 41 (1), 27–50. doi:10.1111/j.1468-2885.2004.tb00302.x
- Lombard, M., Weinstein, L., and Ditton, T. B. (2011). “Measuring telepresence: The validity of the Temple Presence Inventory (TPI) in a gaming context,” in Presented at the 2011 annual conference of the international society for presence research (ISPR), Edinburgh, Scotland.
- Martinez, L., Falvello, V. B., Aviezer, H., and Todorov, A. (2016). Contributions of facial expressions and body language to the rapid perception of dynamic emotions. *Cognition Emot.* 30 (5), 939–952. doi:10.1080/02699931.2015.1035229
- Matsumoto, D., Frank, M. G., and Hwang, H. S. (2012). *Nonverbal communication: Science and applications*. Thousand Oaks, California: Sage Publications.
- McArthur, L. Z., and Baron, R. M. (1983). Toward an ecological theory of social perception. *Psychol. Rev.* 90 (3), 215–238. doi:10.1037/0033-295X.90.3.215
- McGloin, R., Farrar, K. M., and Krcmar, M. (2011). The impact of controller naturalness on spatial presence, gamer enjoyment, and perceived realism in a tennis simulation video game. *Presence Teleoperators Virtual Environ.* 20 (4), 309–324. doi:10.1162/PRES_a_00053
- Minsky, M. (1980). Telepresence. *Omni* 2, 44–52.
- Noël, S., Dumoulin, S., and Lindgaard, G. (2009). “Interpreting human and avatar facial expressions,” in *Human-computer interaction – interact 2009. Lecture Notes in Computer Science* (Springer: Berlin, Heidelberg) 5742. doi:10.1007/978-3-642-03655-2_11

- Normoyle, A., Liu, F., Kapadia, M., Badler, N. I., and Jörg, S. (2013). "The effect of posture and dynamics on the perception of emotion," in *Proceedings of the ACM Symposium on Applied Perception (SAP '13)* (Dublin, Ireland), 91–98. doi:10.1145/2492494.2492500
- Nowak, K. (2001), 2. Philadelphia, PA, 686–710. Defining and differentiating copresence, social presence and presence as transportation *Presence 2001 Conf.*
- Nowak, K. L., and Fox, J. (2018). Avatars and computer-mediated communication: A review of the definitions, uses, and effects of digital representations on communication. *Rev. Commun. Res.* 6, 30–53. doi:10.12840/issn.2255-4165.2018.06.01.015
- Oh, C. S., Bailenson, J. N., and Welch, G. F. (2018). A systematic review of social presence: Definition, antecedents, and implications. *Front. Robotics AI* 5, 114. doi:10.3389/frvt.2018.00114
- Pan, X., and Hamilton, A. F. de 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
- Poeschl, S., and Doering, N. (2013). The German VR simulation realism scale- psychometric construction for virtual reality applications with virtual humans. *Annu. Rev. Cybertherapy Telemedicine* 11, 33–37. doi:10.3233/978-1-61499-282-0-33
- Reynolds, R. M., Novotny, E., Lee, J., Roth, D., and Bente, G. (2019). Ambiguous bodies: The role of displayed arousal in emotion [mis]perception. *J. Nonverbal Behav.* 43 (4), 529–548. doi:10.1007/s10919-019-00312-3
- Rogers, S. L., Broadbent, R., Brown, J., Fraser, A., and Speelman, C. P. (2022). Realistic motion avatars are the future for social interaction in virtual reality. *Front. Virtual Real.* 2. doi:10.3389/frvir.2021.750729
- Schroeder, R. (2006). Being there together and the future of connected presence. *Presence* 15, 438–454. doi:10.1162/pres.15.4.438
- Schroeder, R. (2002). "Copresence and interaction in virtual environments: An overview of the range of issues," in *Proceedings of the 5th international workshop on presence*. Editor F. Ribeiro (São Paulo, Brazil), 274–295.
- Sheridan, T. B. (1992). Musings on telepresence and virtual presence. *Presence Teleoperators Virtual Environ.* 1 (1), 120–126. doi:10.1162/pres.1992.1.1.120
- Shrout, P. E., and Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychol. Bull.* 86 (2), 420–428. doi:10.1037/0033-2909.86.2.420
- Slater, M. (1999). Measuring presence: A response to the witmer and singer presence questionnaire. *Presence* 8 (5), 560–565. doi:10.1162/105474699566477
- Slater, M., Sadagic, A., Usoh, M., and Schroeder, R. (2000). Small-group behavior in a virtual and real environment: A comparative study. *Presence* 9 (1), 37–51. doi:10.1162/105474600566600
- Sun, Y., and Won, A. S. (2021). Despite appearances: Comparing emotion recognition in abstract and humanoid avatars using nonverbal behavior in social virtual reality. *Front. Virtual Real.* 2. doi:10.3389/frvir.2021.694453
- Thie, S., and van Wijk, J. (1998). "A general theory on presence," in *1st international workshop on presence*. Ipswich, Suffolk, UK. <http://www0.cs.ucl.ac.uk/staff/m.slater/BTWorkshop/KPN>.
- Visch, V. T., Goudbeek, M. B., and Mortillaro, M. (2014). Robust anger: Recognition of deteriorated dynamic bodily emotion expressions. *Cognition Emot.* 28 (5), 936–946. doi:10.1080/02699931.2013.865595
- Wagner, H. L. (1993). On measuring performance in category judgment studies of nonverbal behavior. *J. Nonverbal Behav.* 17 (1), 3–28. doi:10.1007/BF00987006
- Watson, D., Clark, L. A., and Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *J. Personality Soc. Psychol.* 54 (6), 1063–1070. doi:10.1037/0022-3514.54.6.1063
- 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
- Worldviz (2021). We build VR labs. Available at: <https://www.worldviz.com>.
- Zebrowitz, L. A., and Collins, M. A. (1997). Accurate social perception at zero acquaintance: The affordances of a gibsonian approach. *Personality Soc. Psychol. Rev.* 1 (3), 204–223. doi:10.1207/s15327957pspr0103_2
- Zebrowitz, L. A. (2002). The affordances of immersive virtual environment technology for studying social affordances. *Psychol. Inq.* 13 (2), 143–145. Retrieved from: <http://www.jstor.org/stable/1449173>.