Interpupillary to Inter-Camera Distance of Video See-Through AR and its Impact on Depth Perception

Franziska Westermeier [*†]
HCI[‡] and PIIS[§] Group, University of Würzburg

Chandni Murmu [*¶]
Clemson University

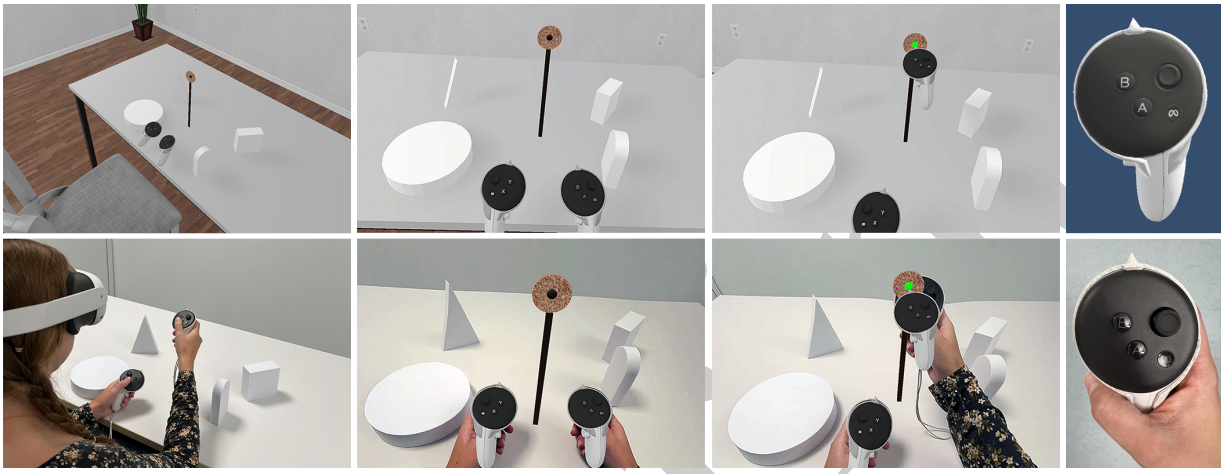
Kristopher Kohm [¶]
Clemson University

Christopher Pagano [¶]
Clemson University

Carolin Wienrich [†]
PIIS[§] Group, University of Würzburg

Sabarish V. Babu [‖¶]
Clemson University

Marc Erich Latoschik [‖†]
HCI[‡] Group, University of Würzburg



(a) The view of the scene with the table and white foam reference objects.

(b) Participant view on the target. As soon as the participant reaches to the target object, the screen turns black.

(c) The calibration phase. After blind reaching, visual feedback is provided when the controller tip is placed in the right position.

(d) The virtual and the physical controller tip.

Figure 1: The virtual and physical scene setup and experimental task in VR (top) and VST AR (bottom).

Abstract

Interpupillary distance (IPD) is the most important parameter for creating a user-specific stereo parallax, which in turn is crucial for correct depth perception. This is why contemporary Head-Mounted Displays (HMDs) offer adjustable lenses to adapt to users' individual IPDs. However, today's Video See-Through Augmented Reality (VST AR) HMDs use fixed camera placements to reconstruct the stereoscopic view of a user's environment. This leads to a potential mismatch between individual IPD settings and the fixed Inter-Camera Distances (ICD), which can lead to perceptual incongruencies, limiting the usability and, potentially, the applicability of VST AR in depth-sensitive use cases. To investigate this incongruency between IPD and ICD, we conducted a 2×3 mixed-factor design user study using a near-field, open-loop reaching task comparing distance judgments of Virtual Reality (VR) and VST AR. We also investigated changes in reaching performance via perceptual calibration by incorporating a feedback phase between pre- and post-phase conditions, with a particular focus on the influence of IPD-ICD differences. Our Linear Mixed Model (LMM) analysis showed a significant difference between VR and VST AR, an effect of IPD-ICD mismatch, and a combined effect of both factors. However, subjective measures showed no effect underlining the subconscious nature of the perception of VST AR. This novel insight and its consequences are discussed specifically for depth perception tasks in AR, eXtended Reality (XR), and potential use cases.

Index Terms: eXtended Reality, Depth Perception, Virtual Reality, Augmented Reality, Video See-Through, Interpupillary Distance, Perception-Action, Perceptuomotor Calibration

1 Introduction

Current Augmented Reality (AR) head-mounted displays (HMDs) either use optical see-through (OST) or video see-through (VST). A specific feature of VST AR is its capability to support various proportions of real-virtual content, enabling almost seamless cross-reality transitions. Accordingly, the development of VST AR has been rapidly evolving, mainly enhancing camera quality, object compositing (including tracking and registration) and rendering to increase passthrough capabilities. These advances open up new use cases and fields of application for VST AR, and many more are evolving. AR applications now target workplace usage and serious use cases in medicine, industry, military, transportation, and other critical areas. While VST AR promises to enhance effectiveness, efficiency, and user experience in these often sensitive new use cases, it becomes necessary to systematically investigate potential technology-related shortcomings in VST AR to ensure its usage is safe, productive, and enjoyable for everyone.

The perception and design challenges of AR were subject to many early works addressing perceptual incongruencies [3, 10, 24].

[*] Shared first-authorship
[†] {franziska.westermeier|carolin.wienrich|marc.latoschik}@uni-wuerzburg.de
[‡] Human-Computer Interaction Group.
[§] Psychology of Intelligent Interactive Systems Group.
[¶] {cmurmu|kckohm|cpagano|sbabu}@clemson.edu
[‖] Shared last-authorship

One key issue in VST AR HMDs is the offset between the passthrough cameras and the human eye. This offset includes both frontal displacement and lateral offset (inter-camera distance, ICD) and varies with each user's interpupillary distance (IPD). The spatial and optical parameters defining stereo parallax differ between camera systems and display configurations. Specifically, while most VST AR HMDs now support adjustment for individual IPDs, fixed ICDs persist due to hardware constraints. Depending on the individual ICD-IPD mismatch, resulting incongruencies between the two stereo views of the virtual and the real-world content could potentially affect depth perception and the overall effectiveness, efficiency, and user experience (UX). Previous depth perception studies have mainly focused on VR and OST AR, identifying systematic underestimations in VR, i.e. the depth compression effect [20, 39]. Even though there is evidence for the same effect in VST AR [37, 44, 52], an IPD-ICD mismatch, potentially leading to and explaining worse results in VST AR compared to VR [52], has not been empirically assessed and evaluated in scientific work so far.

This article examines these potential effects of IPD-ICD mismatches in VST AR HMDs on depth perception. Seventy-two participants from Germany and the United States (US) performed an open-loop blind-reaching task to assess distance judgments in VST AR and VR as a reference. The experiment introduced three phases to investigate learning behavior and potential adaptation to depth misjudgments. Through this study, we answer the research question (RQ): *To what extent does XR mode (VR/VST AR), IPD-ICD difference, and perceptuo-motor calibration affect near-field depth perception in XR?* Our work contributes to the knowledge of the effects of VST AR HMDs and provides guidelines on how incongruent stereoscopy might be counteracted in the future.

2 Related Work

Today, both AR and VR are often grouped together under the term XR (for eXtended Reality). While VR provides immersive displays of solely virtual content (from a purely visual point of view), the main characteristic of AR is the visual combination of real-word physical and computer-generated virtual content [3, 31]. VST AR is a type of AR that builds on the functionality of a VR HMD, adding passthrough cameras on the front face to enable hybrid experiences. These cameras digitize real-world content. Hence, a consistent pixel rasterization and color space application of both virtual and physical content is provided, making it easier to manipulate and integrate content from both sources. However, this approach incorporates a shifted viewpoint of the physical content in VST AR, which can lead to perceptual distortions [3, 10, 42].

2.1 Perceptual Incongruencies in VST AR

In the context of AR, prior works have shown that perceptual incongruencies, such as registration errors, visual mismatches (e.g., resolution, lighting, or color discrepancies), and temporal mismatches (e.g., latency issues) are common [3, 10, 24]. These incongruencies not only affect the perception of plausibility and spatial presence, i.e., the feeling of "being there" [32, 51] but also lead to lower task performance and less accurate depth perception [3, 52]. Such general impacts of lower (sensation and perception) level incongruencies (i.e., a mismatch of the processed and expected information) on higher level performance and XR effects is reflected and predicted by the recent Congruence and Plausibility (CaP) model [25], which we therefore apply as a theoretical basis of the depth-perception study presented here.

2.2 Depth Perception

VR and AR have both been subject to depth perception studies. However, VST AR is underexplored. In VR, a systematic underestimation (aka distance compression) [19, 20, 39] was detected, leading to performance errors such as misjudging how far objects are or impacting users' ability to reach close objects. Studies on OST AR report more accuracy compared to VR and VST AR [1, 19, 38], potentially due to the direct, undistorted view of the environment.

2.2.1 Tasks

There are established tasks to measure participants' depth judgments in VR and AR. However, the nature of tasks possibly leads to distinctive outcomes [34, 35, 48]. Typical tasks include verbal reports [1, 34, 52], where participants estimate egocentric distance during exposure and verbally express their estimates. While this method is straightforward to implement and replicate, it involves not only perceptual processing but also cognitive processing, which can introduce additional bias. Participants must actively recall their knowledge of standardized distance measurement systems, potentially leading to individual variance in the results. Other tasks require motoric action, such as walking to or reaching a certain point at a distance [34, 55], or aligning objects [38, 48]. These tasks follow the concept of perception-action, avoiding cognitive interference. Perception-action tasks can be further divided into open-loop and closed-loop tasks. Closed-loop tasks allow for continuous readjustments by maintaining constant visibility [52]. In contrast, open-loop tasks restrict vision at certain points, preventing visual feedback and thereby fostering the immediate perception-action process. Napieralski et al. [34] directly compared different tasks in the same experimental setup in the near-field distance and showed that a reaching task and verbal reporting lead to different results in VR and real life in the near-field distance.

If users are given the opportunity to engage in manual activity to calibrate to the virtual environment, their distance estimation could become accurate [2]. Perceptuo-motor calibration allows the refinement of task-specific actions through feedback [5] of, e.g., visual or audio-based nature [26]. Through this process, users adapt their actions to the environment without taxing their cognition by attuning to relevant information and adjusting accordingly. Studies in VR have shown that errors arising from the common phenomenon of distance compression can be corrected via feedback from perceptual calibration [23]. Other work shows perceptual calibration can occur in real environments to enhance perception-action [41], given an opportunity to receive feedback and correction. We only found one recent study [15] proving an increased accuracy after perceptuo-motor calibration in VST AR.

2.2.2 Depth Ranges

Studies on depth perception in VR and AR can be categorized by depth range (near-field $<$1.5 m, medium-field 1.5 m - 30 m, and far-field $>$30 m [7, 48]). Vaziri et al. [49] studied depth perception in the medium-field distance using a blind-walking task in VST AR. They tested three conditions with a physical target: (1) an unprocessed real-time view, (2) a line-drawing-style view using Sobel and Canny filters, and (3) a white background with only the target visible, compared to a real-world control condition. While all VST AR conditions led to an underestimation of distances, no significant differences were found between them. The authors suggest that the target object alone provided sufficient depth cues when combined with participants' knowledge of their own eye height and their mental approximation of the ground level. Swan et al. [48] explored depth perception in the far-field distance using an OST AR HMD. They found that depth judgments shifted from underestimation to overestimation at approximately 23 m, suggesting that depth cues change in effectiveness as the distance increases. Strengthening this assumption, Mansour et al. [28] used computer vision to examine the effectiveness of depth cues in different ranges. Their results showed that in the near-field, binocular disparity (stereoscopy) is more effective for depth estimation. In the medium- to far-field, motion parallax becomes a more dominant cue. However, studies in near-field AR are rare.

2.2.3 Impact of IPD Mismatches

An IPD-ICD mismatch was subject to many early conceptual papers focusing on perceptual issues in AR [3, 10, 24], yet without empirical evidence. Some VR studies examined an IPD mismatch (not yet considering the ICD) and the resulting misperception, which can be mathematically calculated. With a deviation of 1 mm in IPD and an object 10.4 m away, displayed at a screen at 80 cm distance, the estimated distance is 12.8 m [10]. Thus, it is not unexpected that the IPD was found to impact several aspects of the XR experience, such as cybersickness [47].

Chakraborty et al. [6] investigated depth perception in VR with an IPD mismatch in the medium-field distance by verbal reporting and blind walking with participants whose IPD was smaller than the HMD IPD setting. Results showed no significant relation between IPD mismatch and misjudgment, even though distances were underestimated in general. Willemsen et al. [56] hypothesized the stereoscopic viewing condition to influence the depth compression effect. However, their different viewing conditions of fixed and adjusted IPD as well as bi-ocular and monocular viewing did not diminish the depth underestimation.

Notably, these studies were conducted in the medium-field distance showing no effect of IPD-mismatch [6, 56]. Studies combining *near-field* depth perception with an IPD mismatch are rare, not even talking about the IPD-*ICD* mismatch in VST AR. Compared to medium-field, the likeliness for effects of a mismatching IPD in the near-field is higher, because if the observed object is nearer, the disparity is higher [57] having a greater impact to depth misjudgments. When the object moves farther away, the convergence point of the eyes moves with it until the eye directions are almost parallel.

3 Present Approach

3.1 Hypotheses

Most empirical studies examined incongruent stereoscopy in the medium-field distance with no effect on depth perception [6, 56]. However, according to geometric calculations of the disparity [57], stereoscopy becomes a more important cue in the near-field distance [28, 39]. In addition, most studies focus on VR, leaving many open questions to VST AR including variance given by the ICD. We need to consider that VST AR has an even more complex setup by the desiderate to merge physical and virtual worlds together potentially causing perceptual incongruencies [3, 10, 24, 25]. Thus, we would expect worse results for VST AR (H1.1) than for VR, where the composition of content is visually congruent [25, 52]. Additionally, we expect that a higher IPD-ICD mismatch (i.e. deviation of IPD from the ICD) leads to lower accuracy (H1.2).

H1.1 Depth misjudgment will be higher in VST AR than in VR.

H1.2 Depth misjudgments in VST AR will be higher the higher the deviation between IPD and ICD.

Perceptual calibration for near-field depth perception has been shown to effectively reduce perceptual errors in factors such as depth, size and reach boundary estimation in VR [2, 8, 11, 12] raising potential to deploy in, e.g., training scenarios with precise interaction. Gagnon et al. [15] found an effect by action calibration in VST AR. Calibration in real environments has proved to be effective [41]. Therefore, we expect performance increases in VR [13] as well as in AR [15]. If perceptuo-motor calibration can potentially overcome depth misperception in VST AR viewing due to IPD-ICD mismatch, then calibration can offer a technical solution to enhance depth perception in VST AR in contemporary XR HMDs.

H2.1 Depth misjudgment in VR will be lower after a calibration phase.

H2.2 Depth misjudgment in VST AR will be lower after a calibration phase.

Following the CaP model's assumptions [25], VST AR inheres perceptual incongruencies, leading to conflicting visual cues [3]. Hence, we argue that these perceptual incongruencies (including an IPD-ICD mismatch) might also influence spatial presence in a similar way as the depth judgments.

H3.1 Spatial presence will be higher in VR than in VST AR.

3.2 Study Design

To answer the hypotheses, we apply a 2×3 mixed design. We define the XR mode as between factor with VR and AR (which is realized in VST AR, but we refer to it as AR in the method description and results for easier readability) as the two conditions, and the phase as within factor with the three conditions: pre-, calibration, and post-phase. These phases run in a fixed order, each including 30 trials. In the calibration phase, participants get feedback on their depth judgments. Another factor is the participants' IPD values that we measure and adjust the HMD inner lenses to. Hence, the view on virtual content shall always be congruent with the participant's IPD. However, the passthrough view in VST AR is affected by an IPD-ICD discrepancy.

4 Method

4.1 Participants

We recruited 75 participants distributed in two different locations in Germany and the US. Due to technical issues with recording the data, three participants were excluded, resulting in 72 datasets to evaluate (50 from Germany and 22 from the US). Participants were randomly assigned to the conditions, resulting in 35 participants in the VST AR condition and 37 participants in the VR condition. The study received ethical approval from both Institutional Review Boards (Ethics Committee of the Institute Human-Computer-Media, University of Würzburg, IRB Clemson University). Participants (33 females, 39 males) were aged 20 to 64 years ($M = 29.67$, $SD = 10.68$). Eligibility criteria required participants to have normal or corrected-to-normal visual acuity. Participants' IPDs ranged from 58 to 71 mm with a mean of $M = 64.0\ mm$ ($SD = 3.3\ mm$)(see Fig. 2).

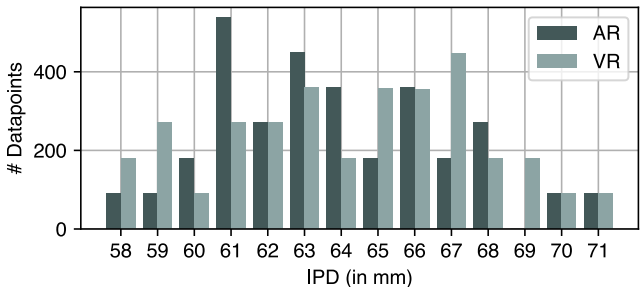


Figure 2: Distribution of IPD across all participants.

To determine the necessary sample size, an apriori power analysis was performed using G*Power (3.1.9.7) [14]. For the analysis, we defined 3 bins for different IPDs. Binning followed the mean IPD of $M = 63.36\ mm$ from Dodgson [9], the boundaries of adjustability in the Meta Quest 3 (58 mm to 70 mm) and the aim to make bins as equidistant as possible. Thus, boundaries are set at 61.5 mm and 65.5 mm. In the later analysis, we treated the IPD as a continuous variable. Based on an effect size of 0.20, an α error probability of 0.05, a power (1-β error probability) of 0.95, two between groups, and 90 measurements, the required sample size was

calculated to be 12 participants per condition. Accordingly, 72 participants were recruited (12 participants per IPD bin across 2 XR modes). An Analysis of Variance (ANOVA) revealed no significant difference between the IPDs present in the VR condition compared to the IPDs in AR ($F(1,70) = 0.005$, $p = .941$).

4.2 Apparatus

Hardware The Meta Quest 3 (v67/v68) was used at 90 Hz and connected to the computer with an Oculus Link cable. The computer in Germany was equipped with an Nvidia Geforce RTX 3080 GPU and an Intel(R) Core(TM) i9-11900K CPU with 64 GB RAM. In the US, an Nvidia GeForce RTX 3070 and an Intel(R) Core(TM) i5-10500 with 32 GB RAM were used. We evaluated the frames per second (fps) to ensure similar performances. Results showed comparability between both machines with mean fps of $M = 89.90\ Hz$, ($SD = 0.60\ Hz$) overall, $M = 89.91\ Hz$, ($SD = 0.64\ Hz$) for Germany, and $M = 89.87\ Hz$, ($SD = 0.52\ Hz$) for the US.

Controller Tip Since the controller's volume is too large for precision pointing in 3D, we designed and 3D-printed a controller tip at the front face of the controller that was used as the reference point for the distance measurement (see Fig. 1d, bottom). The 3D model was also added to the virtual controller to provide the same virtual representation (see Fig. 1d, top). To get accustomed to the position of the controller tip, we conducted some pointing and touching tasks (e.g., "Please touch the tip of your index finger of your non-dominant hand with the controller tip") before the exposure.

Software We used the Unity engine (2021.3.27) and the Realitystack I/O framework [21] to support the HMD and controller interaction. For data preparation and analysis, Python (3.8.19) and R (4.4.1) were used. For the calculation of Linear Mixed Models (LMMs), we further used the 'buildmer' [50], the 'lme4' [4] and the 'performance' package [27]). The figures for these models were created using the 'ggplot2' package [53] in R.

4.3 Depth Judgment Task

We assessed participants' depth judgments using an open-loop task in three phases: the pre-phase, the calibration phase, and the post-phase. We used open-loop reaching because it is more accurate, less variable, and less influenced by cognitive factors than other approaches [36]. The task procedure is consistent in both the *pre-phase* and *post-phase*. The environment is first displayed to the participant — either in VR or AR. Two seconds later, the target object appears, represented by a textured disk with a marked center on a stick (see Fig. 1b). Participants reach forward with the controller of their dominant hand. As the complete controller mesh moves out of a predefined cylindrical boundary (0.18 m radius, placed at the table's edge at chair height), the screen turns black, requiring participants to judge the position of the target object without visual feedback. By allowing the participants to initiate the black screen themselves, visual feedback is immediately limited once they have an understanding of their spatial surroundings as they move during the open-loop task [36]. Participants confirm their judgments by pressing and holding the trigger button on their dominant hand while pressing the trigger on their non-dominant hand once.

The *calibration phase* uses a similar open-loop task as in the pre- and post-phases but with an additional step for readjustment and feedback. After making their initial judgment, participants' vision is immediately restored, allowing them to see the difference between their judgment and the actual target position. They then readjust the controller to align with the target object. When the controller touches the center of the target, the object turns green, providing visual feedback (see Fig. 1c). Participants confirm their adjustment again. The screen then turns black, and they return the controller to the resting position before the next trial begins.

The target objects are placed in predefined positions based on each participant's maximum arm reach. In the pre- and post-phases, six different target distances, are used, ranging from 30% to 80% of the participant's arm reach in increments of 10%. These six distances are repeated five times in a randomized order, resulting in 30 trials per phase [46]. During the calibration phase, target distances are set between 35% and 85% of arm reach to prevent participants from calibrating to specific distances used in the pre- and post-phases. All positions are defined at the start of the experiment, using the HMD head position as the origin, with targets aligned along the z-axis. Beforehand, a room calibration is conducted to align the virtual and physical space and ensure that participants are directed towards the z-direction. We designated two corners of the table as reference points and registered the tip of the physical controller at these corners. This allowed the virtual controllers to be positioned in the same relation to the virtual table (similar to [52]).

To engage participants with the physical/virtual environment and to provoke a switch from a focus on the passthrough view to the virtual content in the AR condition, we add reference objects to the table as visual anchors (see Fig. 1a). We apply a waiting time of two seconds for participants to inspect the environment before the target object appears, and participants switch their focus to the target object.

4.4 Procedure

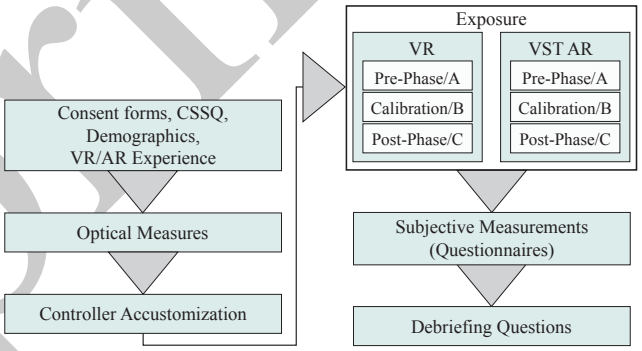


Figure 3: Experiment procedure

The procedure can be seen in Fig. 3. It starts with signing consent forms and a cybersickness screening questionnaire (CSSQ). We assess demographic data and media usage, as well as previous VR/AR experience. In the next step, we take some optical measures. Then, participants get accustomed to the controller. Participants then put on the HMD, and the arm length is measured. Participants stretch their arms forward, holding the controller. Then, we calculate the forward distance of the HMD to the orthogonal plane in which the controller tip is placed. After that, the room is calibrated, and participants listen to audio instructions. They start the three phases subsequently, either in VR or AR, without breaks in between. Each phase consists of 30 trials. Participants are introduced to each phase with an audio explanation and a black screen displaying the next phase (Phase A/B/C). They are told to move their body as little as possible, except for their arms. After phase C, participants take off the HMD and answer some questionnaires and debriefing questions.

4.5 Measures

4.5.1 Optical Measures

Color vision was assessed using the Ishihara Color Vision test [18]. To control for the stereo acuity of participants, which is essential for depth perception, we conducted a stereo acuity test with the Fly-S test [29]. To measure the IPD, we used a mirror measurement as described by Willemsen et al. [56] and conducted

three repetitions of this measure. We used the apps Eyemeasure[1] and GlassesOn[2] in addition, to factor out variance of participant-specific (in)accuracies. The correlation between the mean of manual measurements and the app measurement amounts to 0.767 and 0.771 for the two apps used, indicating a strong positive relationship. We decided on the measurement by app for further usage and analysis.

4.5.2 Objective Measures

We track the head position at the center point between both eyes, the position of the actual target, and the perceived target position (i.e., where participants estimate the target object to be). To set them in relation, we calculated the distances in millimeter (mm) in the z-direction from the head to the actual and to the perceived target, respectively. We included this perceived distance and calculated the signed residual error ($SRE = perceived - actual$) and the signed proportional error ($SPE = (perceived - actual)/actual * 100$) for our evaluation.

We tracked the time needed for each trial in milliseconds (ms). The time started as soon as the screen turned black and stopped with the confirmation by controller input.

Our analyses of the objective measures were performed using LMMs. LMMs incorporate both *fixed* and *random* effects, allowing researchers to see how predictors of various data types influence the outcome variable as well as variability such as differences between participants [30]. They are more robust than ANOVAs, and especially useful for repeated or unbalanced measurements (e.g., IPDs, see Fig. 2). With the use of LMMs, we could treat the IPD as a continuous variable (rather than defining categorical bins for ranges of IPDs as we did for the initial power analysis) for a more detailed analysis. We checked violations of assumptions such as multicollinearity (see the `check_model` function in the 'performance' R package [27]). Any predictors that violated these assumptions were removed, and the models were re-checked for theoretical soundness and best fit. For each LMM model, we show both marginal (m.r^2) and conditional r^2 (c.r^2) values. The marginal r^2 values provide the effect size of each predictor in the model, and the conditional r^2 values show the overall ability of the model to explain variance in the outcome variable [33].

4.5.3 Subjective Measures

Participants completed the Igroup Presence Questionnaire (IPQ) [45] and the Spatial Presence Experience Scale (SPES) [17] to assess presence and spatial presence, respectively. The IPQ consists of 14 items across three subscales: *spatial presence*, *involvement*, and *experienced realism*, rated on a 7-point Likert scale (1 = strongly disagree, 7 = strongly agree). For this study, the questionnaire was adapted for AR by replacing terms like "virtual environment" with "shown environment." The SPES, which can be applied to various media formats and thus, is not restricted to VR, consists of eight items divided into two subscales *self-location* and *possible actions*, with responses ranging from 1 ("I do not agree at all") to 5 ("I fully agree").

To understand the participants' frame of reference as control measure — whether they perceived the content as more real or virtual [54] — we asked them to rate their perception of objects, the environment, interactions, the scenario, and the overall experience using a continuous slider from -1 (completely real) to 1 (completely virtual).

Since we are extending the controller with a tip, we were interested if participants felt an altered embodiment. Hence, we used the existing Virtual Embodiment Questionnaire (VEQ) by Roth and Latoschik [43] and adapted the questions. They are answered on a 7-point Likert scale (1 = strongly disagree, 7 = fully agree). The items are assigned to three dimensions *ownership* ("It felt like the controller belonged to me", "It felt like the controller was part of my body"), *agency* ("I felt like I was controlling the movements of the controller", "The movements of the controller were in sync with my own movements", "The controller accurately followed my hand movements", "I felt like I was causing the movements of the controller", and *change* ("The controller facilitated me to reach the targets in my near field space", "I felt like the controller changed my reaching behavior").

Task load was measured using the NASA Task Load Index (NASA-TLX) [16]. Participants rated their workload on a continuous scale from 0 to 20.

VR sickness was monitored using the Virtual Reality Sickness Questionnaire (VRSQ) [22], which participants completed on a 4-point Likert scale (0 = none, 3 = severe). Additionally, a pre-experiment cybersickness screening (CSSQ) was conducted to assess participants' susceptibility to motion sickness, with questions such as "I feel sick when I go backwards by train" to identify and possibly exclude highly susceptible participants.

We conducted a textual debriefing interview, asking for qualitative feedback on participants' experiences, perceptions, and any additional thoughts or suggestions they had regarding the study.

5 Results

5.1 Objective Measures

Our LMM models used several predictors (see Tab. 1) and their interactions for outcome estimation. Before creating each of the LMMs, we removed outliers for each of the model's dependent variables using a z-score of three in the `check_outliers`() from the 'performance' R package [27]. Out of 6,467 total data points, 0 data points were removed for the estimated distances in the first model, 280 data points were removed for the residual error in the second model, and 305 data points were removed for the proportional error.

Table 1: Table of Predictors

Predictor	Context/Formula
Actual Distance	Target object position - current head position (center between both eyes) in z-direction (mm)
XR Mode	VR and AR *(Baseline: VR. Interpret as: Impact on outcome when condition changes from VR to AR)*
IPD-ICD Difference	Participant's IPD - Quest 3's fixed ICD (mm) *(−): IPD <ICD, (+): IPD>ICD*
Trial Phase	Pre-phase, Calibration phase, and Post-phase *(Baseline: Pre-phase. Trial phase has two transitions with identical marginal r^2 values: pre to calibration and pre to post)*
Trial Number	Iterations in each phase: 30, Total: 30 × 3 = 90
Trial Duration	Duration of each trial (ms)

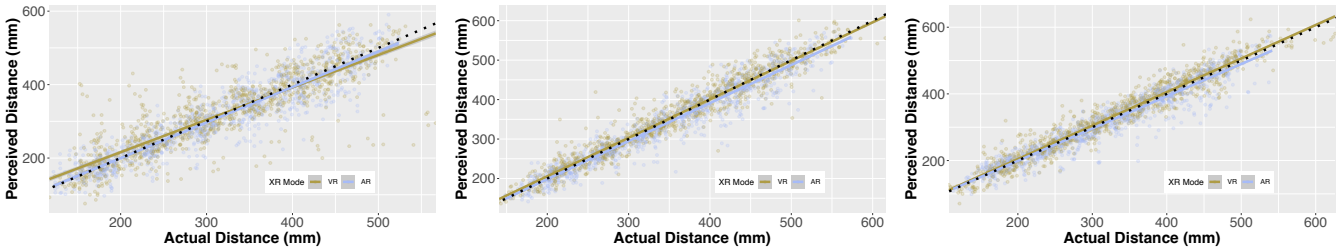
5.1.1 Model 1: Perceived vs. Actual Distance

Model 1 predicts the *perceived distance* (in mm). It is the difference in the z-direction between the target object's perceived position and the participant's head position (center between the eyes). This facilitates direct comparison of machine-processed distance in standardized units. All predictors (see Tab. 1) were used as fixed effects and participant ID as a random effect. The model had strong marginal and conditional explanatory power (m.$r^2 = 0.89$, c.$r^2 = 0.93$) for the fixed effects (see Tab. 2 for parameter values and Fig. 4 for the perceived vs. actual distances in the XR mode for all trial phases).

The model's intercept (15.06 mm) was significant. Except trial duration and post-phase (compared to pre-phase) all other predictors were statistically significant. The actual distance increased per-

[1] https://apps.apple.com/us/app/eyemeasure/

[2] https://play.google.com/store/apps/details?id=com.sixoversix.copyglass



(a) Pre-phase (b) Calibration phase (c) Post-phase

Figure 4: These graphs show the participants' distance judgments in the three phases separated by both XR modes against the actual distances that they were trying to reach. Perfect reaching would result in a slope of one and would line up with the black dotted line in each graph. Estimations were closer to the actual distances in the calibration and post-calibration phases, showing the impact of perceptual calibration.

ceived distance by 0.97 mm for every 1 mm increase in the actual position, suggesting a consistent underestimation (see Fig. 4). Switching from VR to AR increased underestimation by 13.99 mm. The IPD-ICD difference further decreased perceived distance by 2.57 mm for every 1 mm increase in the signed IPD-ICD difference. The interaction between the XR mode and the IPD-ICD difference increased perceived distance by 3.46 mm, suggesting that the IPD-ICD difference moderates the impact of the XR mode. The trial number also moderated the effect of the IPD-ICD difference, reducing underestimation slightly. Other significant interactions included trial duration moderating both the XR mode and the IPD-ICD difference, as well as trial phase (pre-to-calibration) moderating the effect of the IPD-ICD difference.

The actual distance had the largest impact on perceived distance, contributing significantly to the model's explanatory power. This indicates that participants consistently underestimated distances, and the underestimation was influenced by interactions with other factors.

Table 2: Perceived vs. Actual Distance.

Fixed Effect	Beta [95% CI]		t(6451)	p	$m.r^2$
(Intercept)	15.06	[7.76, 22.37]	4.04	**<.001**	–
Actual distance (mm)	0.97	[0.96, 0.97]	262.01	**<.001**	.86
XR Mode [AR]	-13.99	[-23.54, -4.44]	-2.87	**.005**	.004
IPD-ICD Diff. (mm)	-2.57	[-4.49, -0.64]	-2.61	**.011**	.0006
Trial number	0.11	[0.02, 0.19]	2.53	**.011**	<.0001
Trial dur. (ms)	0.00	[-0.001, 0.001]	0.21	.832	-2.34e-12
Phase [Cal.]	-1.85	[-3.61, -0.09]	-2.06	**.039**	.00007
Phase [Post]	0.48	[-1.27, 2.23]	0.54	.592	.00007
XR Mode [AR]:IPD-ICD	3.46	[0.66, 6.26]	2.43	**.018**	.0021
IPD-ICD:Trial num.	-0.03	[-0.05, -0.003]	-2.16	**.031**	<.00001
XR Mode [AR]:Trial dur.	0.002	[0.001, 0.004]	2.67	**.008**	.0002
IPD-ICD:Trial dur.	-0.00	[-0.001, -0.0002]	-3.83	**<.001**	-.0003
IPD-ICD:Phase [Cal.]	2.37	[1.84, 2.89]	8.81	**<.001**	.0010
IPD-ICD:Phase [Post]	2.03	[1.51, 2.56]	7.56	**<.001**	.0010

Cal. = Calibration, Diff. = Difference, Dur. = Duration.

5.1.2 Model 2: Signed Residual Error (SRE)

The signed residual error (SRE) is the difference between the perceived and the actual target distance ($perceived - actual$). It captures the magnitude and direction of errors (underestimation when negative/overestimation when positive), revealing behavioral patterns. Actual distance was used in calculating SRE, other predictors were used as fixed effects and participant ID as random effects. Target distance was used as additional fixed effect to observe the impact of target position on SRE. The model revealed power of m. $r^2 = 0.05$ and c. $r^2 = 0.34$ (see Tab. 3 for parameter values and Fig. 5 for the interaction effects of IPD-ICD difference and XR mode on SRE).

The intercept was significant at 6.09 mm. Except IPD-ICD difference and post-phase (compared to pre-phase) all other predictors were statistically significant. Transitioning from VR to AR significantly decreased the SRE, increasing underestimation by 7.53 mm. Target distance also significantly decreased the SRE, with each unit increase in target distance resulting in an 8.41 mm increase in underestimation. Similarly, transitioning from the pre- to calibration phase significantly decreased the SRE by 2.39 mm, leading to further underestimation. Both trial duration and trial number significantly increased the SRE (0.001 mm/ms of trial duration and 0.08 mm per trial) indicating that participants became more accurate the longer and more frequently they performed the task.

The interaction between XR mode and IPD-ICD difference significantly increased the SRE, suggesting that the IPD-ICD difference moderated the effect of XR mode. 1 mm increase in the signed IPD-ICD difference when changing from VR to AR resulted in a 2.43 mm increase in the SRE, leading to overestimation. Similarly, the IPD-ICD difference moderated the transitional effect of pre- to calibration phase, causing an overestimation of 1.04 mm, and pre- to post-phase, leading to an overestimation of 0.69 mm. IPD-ICD difference also moderated the effects of trial duration and trial number, increasing overall underestimation when interacting with each of them. Since the main effects of both the IPD-ICD difference and the trial phase were not significant, these interactions warrant further investigation.

The interaction between XR mode and IPD-ICD difference had the largest standardized beta value in the model, while XR mode alone had the largest m.r^2. Together, this interaction term and the main effect of transitioning from VR to AR had the greatest impact on the SRE. Notably, the standardized beta value for the IPD-ICD difference was comparable to that of the XR mode, indicating a substantial impact on the SRE, though its m.r^2 was smaller than that of the XR mode and its interaction term.

Table 3: Signed Residual Error (SRE).

Fixed Effect	Beta [95% CI]		t(6172)	p	m. r^2
(Intercept)	6.09	[0.60, 11.58]	2.18	**.031**	–
XR Mode [AR]	-7.53	[-14.49, -0.57]	-2.12	**.037**	.034
IPD-ICD Diff. (mm)	-1.06	[-2.54, 0.41]	-1.41	.161	.004
Target distance (%)	-8.41	[-11.68, -5.13]	-5.03	**<.001**	.003
Phase [Cal.]	-2.39	[-3.78, -1.00]	-3.36	**<.001**	.002
Phase [Post]	0.36	[-1.03, 1.74]	0.51	.613	.002
Trial dur. (ms)	0.0014	[0.0007, 0.0021]	4.08	**<.001**	.003
Trial no.	0.08	[0.01, 0.14]	2.30	**.021**	.001
XR Mode [AR]:IPD-ICD	2.43	[0.31, 4.55]	2.24	**.028**	.021
IPD-ICD:Phase [Cal.]	1.04	[0.62, 1.46]	4.84	**<.001**	.003
IPD-ICD:Phase [Post]	0.69	[0.27, 1.12]	3.22	**.001**	.003
IPD-ICD:Trial dur. (ms)	-0.0002	[-0.0004, -0.00006]	-2.68	**.007**	.001
IPD-ICD:Trial no.	-0.03	[-0.05, -0.01]	-3.19	**.001**	.001

Cal. = Calibration, Diff. = Difference, Dur. = Duration.

5.1.3 Model 3: Signed Proportional Error (SPE)

Model 3 predicts the signed proportional error (SPE), calculated as (perceived − actual)/actual × 100. It accounts for scaling effects, reflecting relative deviation, as humans perceive distance proportionally, regardless of units, allowing for better comparability with related work. Predictors (see Tab. 1) were used as fixed effects and participant ID as random effects. The model revealed a $m.r^2 = 0.04$ and $c.r^2 = 0.31$ (see Tab. 4 for parameter values and Fig. 6 for the impact on SPE when IPD-ICD difference and XR mode interact).

Apart from the model's intercept and post-phase (compared to pre-phase), all the other effects were statistically significant. Transitioning from VR to AR condition significantly decreased the SPE, indicating that the AR condition increased underestimation by 2.40%. Similarly, the IPD-ICD difference significantly decreased the SPE: for every 1 mm increase of the signed IPD-ICD difference, underestimation increased by 0.53%. Transitioning from the pre- to calibration phase also significantly decreased the SPE, increasing underestimation by 0.50%. Additionally, both trial number and trial duration significantly increased the SPE by 0.03% per trial and 0.0003% per ms of trial duration, respectively. These increases in SPE values correspond to reductions in underestimation; hence, participants became more accurate with more trials and as they took more time.

The interaction between XR mode and IPD-ICD difference indicates that changing conditions from VR to AR moderated the additional underestimation caused by the main effect of IPD-ICD difference, reducing its impact by 0.73%. This effect is larger than the main effect of IPD-ICD difference alone, thereby reducing underestimation overall. Additionally, the interaction between IPD-ICD difference and trial phases (pre-to-calibration and pre-to-post) significantly increased proportional error. The IPD-ICD difference moderated the effect of the pre-to-calibration phase by reducing the increase in underestimation and further moderated the pre-to-post phase by amplifying the decrease in underestimation. The IPD-ICD difference also moderated the effect of the trial number by slightly increasing underestimation.

The interaction between the XR mode and the IPD-ICD difference had the largest standardized beta value in the model, whereas the XR mode alone had the largest m.r^2 value. Together, these terms had the greatest effect on the SPE, similar to the SRE model.

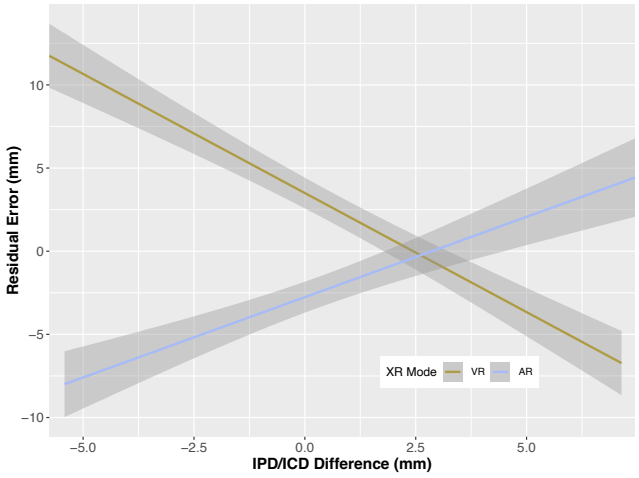


Figure 5: This graph shows the relationship between the SRE and the IPD-ICD difference, separated by the XR mode. It demonstrates the significant interaction between the IPD-ICD difference and the XR mode, as well as the main effect of each.

Table 4: Signed Proportional Error (SPE).

Fixed Effect	Beta [95% CI]		t(6149)	p	m.r^2
(Intercept)	0.24	[-1.36, 1.86]	0.30	.762	–
XR Mode [AR]	-2.40	[-4.54, -0.26]	-2.20	**.031**	.033
IPD-ICD Diff. (mm)	-0.53	[-0.97, -0.09]	-2.34	**.021**	.003
Phase [Cal.]	-0.50	[-0.95, -0.05]	-2.20	**.028**	.002
Phase [Post]	0.41	[-0.04, 0.86]	1.79	.073	.002
Trial no.	0.03	[0.009, 0.05]	2.85	**.004**	.001
Trial dur. (ms)	0.0003	[0.0001, 0.0005]	2.91	**.003**	.002
XR Mode [AR]:IPD-ICD	0.73	[0.08, 1.38]	2.20	**.031**	.019
IPD-ICD:Phase [Cal.]	0.32	[0.18, 0.46]	4.59	**<.001**	.003
IPD-ICD:Phase [Post]	0.20	[0.06, 0.34]	2.89	**.003**	.003
IPD-ICD:Trial no.	-0.006	[-0.01, -0.00009]	-1.99	**.046**	.0004

Cal. = Calibration, Diff. = Difference, Dur. = Duration.

5.2 Subjective Measures

Subjective measures were evaluated using a one-way ANOVA. Homogeneity (Levene's test) was given for all dependent variables; data being normal-distributed was not always true (Shapiro-Wilk test). Hence, we cross-checked our ANOVA results with a non-parametric Kruskal-Wallis Test, leading to consistent (non)significance patterns. We did not find significant values in the IPQ ($F(1,70) = 0.076, p = .783, \eta_p^2 = .001$) or any of the IPQ subscales (spatial presence: $F(1,70) = 1.844, p = .179, \eta_p^2 = .026$); neither did we find significant deviations between VR and AR in the SPES ($F(1,70) = 0.010, p = .920, \eta_p^2 < .001$) or any of its subscales. When we asked participants about their reference frame, two items showed significance. Participants perceived the environment in the AR condition as more real ($F(1,70) = 8.529, p = .004, \eta_p^2 = .109$; VR $M = 0.09, SD = 0.55$; AR $M = -0.28, SD = 0.53$) and the interaction in VR as more real ($F(1,70) = 5.824, p = .018, \eta_p^2 = .015$; VR $M = -0.17, SD = 0.44$; AR $M = 0.11, SD = 0.54$).

5.2.1 Control Measures

The control measures did not show significant results. The mean values of the adapted VEQ were high ($M = 5.10$, $SD = 0.74$). The NASA TLX values showed medium to low ratings ($M = 6.29$,

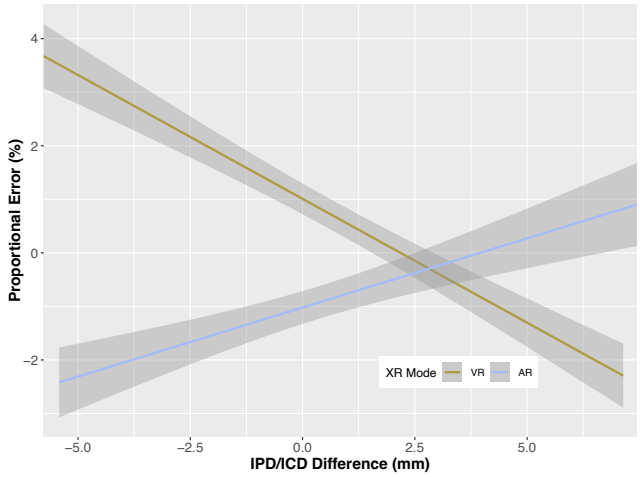


Figure 6: This graph shows the relationship between the SPE and the IPD-ICD difference, separated by the XR mode. This graph is similar to the graph showing the SRE, both showing the main effects of the IPD-ICD difference, XR mode, and their interaction. It is worth noting that the IPD-ICD difference, at which the two XR modes have the same SPE, is a different difference-value than the one at which the XR modes have the same SRE.

$SD = 2.96$) with no indications of significance. There were neither significant differences nor abnormalities in the VRSQ ratings ($M = 8.53$, $SD = 11.76$). Hence, we could exclude confounds given by potential VR sickness symptoms and ensure comparability of task difficulty between VR and AR.

Debriefing Interview Participants were asked to compare their performance between the pre- and post-phase. Seventeen participants reported no perceived change in their performance, while two were unsure. Additionally, eleven participants mentioned a perceived change in their performance but did not specify whether it was positive or negative. Among those who did specify, 31 participants reported a positive change in performance (higher accuracy, more confidence), while five participants reported a negative change (lower accuracy, more frustration). Six participants did not respond to that question. When we asked participants about the effectiveness of the calibration phase, 23 participants mentioned that they actively applied a correction in the post-phase, which they learned in the calibration phase. Thirteen additional participants stated that the calibration phase contributed to better judgments, and nine participants explicitly stated that the calibration phase had no influence on their judgment in the post-phase.

6 Discussion

In the beginning, we posed the RQ: *To what extent XR mode (VR/AR), IPD-ICD difference, and perceptuo-motor calibration affects near-field depth perception in XR?* Our RQ was operationalized using several hypotheses stated in Sec. 3. In the following discussion, we evaluate these hypotheses in the process of answering the overarching RQ of the study.

6.1 Depth Judgments in VR versus VST AR

H1.1 hypothesized that "Depth misjudgment will be higher in VST AR than in VR." This was supported by model 1 (see Tab. 2) where misjudgment is evident by the $\approx$ 14 mm shorter perceived distance in VST AR when compared to VR. Examining measures of accuracy revealed that participants' signed residual error (SRE, *perceived* − *actual*, model 2) increased by $\approx$ 8 mm in VST AR overall as compared to VR. Also, examining the proportion of the signed error by target distance as a percentage (or proportional error, SPE, model 3) revealed an increase by 2.40% in VST AR for each mm of actual distance, as compared to VR.

Prior work has already described the bandwidth of perceptual issues present in VST AR HMDs [3, 10, 24], whereas the IPD-ICD mismatch is only one factor leading to a distorted viewing condition. Also other factors, such as the camera-eye offset of the passthrough cameras in the z-direction or differences in lens distortions, let the view appear incongruent, challenging the user to interact properly. Visual mismatches, such as differences in color or contrast, can also be a factor to convey conflicting depth information. The results support Azuma's [3] elaboration on how contradicting visual-visual stimuli (VST AR, physical and virtual stimuli) are misperceived way more easily than contradicting stimuli across multiple senses (VR, visual sense and, e.g., vestibular sense).

6.2 Effects of IPD-ICD Difference

H1.2 posited that "Depth misjudgments in VST AR will be higher the higher the deviation between IPD and ICD." This hypothesis is supported by our data across several measures. In the LMM model of the relationship between actual and perceived distance (model 1), the IPD-ICD difference variable was a significant predictor of perceived distance. Each unit of increasing the signed IPD-ICD difference results in a decrease of $\approx$ 3 mm of perceived distance, resulting in an equivalent amount of underestimation in the perceived distance. Similar results were found with regard to the effects of IPD-ICD difference on the accuracy scores (SRE and SPE).

For H1.1, we could identify a difference in perceived depth, residual, and proportional errors between VST AR and VR. Notably, these results are also explained or moderated by the IPD-ICD difference. Thus, as compared to the prior work on near-field depth perception in VR and AR, our data shows that IPD-ICD deviation is a plausible factor for why VST AR has more error than VR viewing in near-field depth perception [52]. The discrepancy caused by viewing the foreground virtual target object with the IPD stereo disparity (adjusted to users' IPD), juxtaposed against the physical environment viewed through the passthrough with the ICD stereo disparity, may amplify to a poorer depth perception in VST AR as compared to VR. This underlines previous research that assumed that the fixed ICD creates perceptual issues [3, 10, 24, 42] in VST AR. However, this has not been proven so far.

6.3 Learning Effects by Perceptuo-Motor Calibration

H2.1: "Depth misjudgment in VR will be lower after a calibration phase." and **H2.2**: "Depth misjudgment in VST AR will be lower after a calibration phase." are partially supported because there was no significant interaction effect between XR mode and the trial phase. However, we did find evidence that the underestimation of distance improved significantly in both conditions and in the trial phase. From model 1, we observed a significant effect in the transition from pre- to calibration phases. Though the perceived distance decreased from the pre- to calibration phase, the IPD-ICD difference moderated this effect, increasing perceived distance overall and reducing underestimation. While no significant effects were found when moving from the pre- to the post-phase, there was a significant interaction between the IPD-ICD difference and the transition from pre- to post-phase. This suggests that the IPD-ICD difference can impact the effect of calibration and requires further investigation. In model 2, a significant main effect was observed during the transition from pre- to calibration, which decreased signed error and, consequently, increased underestimation. However, the interaction between the IPD-ICD difference and the trial phase (from pre to calibration phase) counteracted the increase in underestimation. Despite this, there was an overall rise in underestimation, likely explained by the significant intercept in model 2. Model 3 revealed a significant effect for transition from pre- to calibration phase where the proportional error value decreased, meaning underestimation increased. Significant interactions between IPD-ICD differences and both transitions moderated the main effects, reducing the severity of underestimation overall.

Overall, as the main effect of trial phase — pre- to post — was not significant, we did not find evidence that the calibration task performed in between pre- and post-phases significantly reduced perceptual error in the two XR modes. We expected an impact of calibration at least in the VR condition given past research [2, 5, 23, 26], therefore we provide a few reasons why this may not have been the case in this study. It is possible that the task was easier than previous studies, leaving insufficient room for improvement. Though Fig. 4a for the perceived distance in the pre-phase shows more underestimation than in the calibration or post-phase, perceived distances were still fairly accurate. Also, fatigue from the length of the task may have reduced participants' performance in the post-phase. Given the debriefing interviews, it is worth noting that many participants perceived a positive change in their performance from the calibration phase. Also, the significant interactions between the IPD-ICD difference and calibration (seen in the trial phase going from pre- to post) should be investigated further in the VST AR because this difference may be impacting participants' ability to calibrate as well.

6.4 Subjective Findings

We did not find any effects on presence or spatial presence. Hence, **H3.1**: "Spatial presence will be higher in VR than in VST AR"

is not supported. However, the nature of the task restricted users' head movements, explicitly avoiding motion parallax as a strong promoter for spatial presence. Furthermore, the task description strictly determined how to act and thus, there was little scope for exploring the environment and differentiation between the VR and AR condition. In addition, one subscale of the SPES focuses on possible actions and asks for hypothetic interactability with objects. Since participants only interacted with one object, this measurement might not be sensitive enough to capture effects.

The environment item of the reference frame questions revealed significance. Participants perceived the VST AR environment as more real than the VR environment. Notably, all other items showed inverted ratings. The objects, the scenario, the interaction, and the experience were rated as more real in VR. We assume that in the VST AR condition, participants used the environment as the frame of reference to judge the experience [54]. Thus, they were more strict in VST AR when judging the other items, as the contrast between the environment and the virtual content is too high. In VR, participants experienced one congruent scenario, which did not lead to a drop in the perception of realism in any of these aspects.

According to objective measures, participants' ability to judge spatial relations towards a target object was clearly reduced in VST AR. However, subjective measures, particularly those related to spatial presence, did not show significant differences when compared to our objective measures. This highlights a discrepancy between objective performance and subjective assessments. We assume that questionnaire-based assessments may be too insensitive to detect subtle variations in depth perception. Additionally, completing questionnaires requires cognitive resources, which might introduce biases that favor VST AR. These cognitive influences could neglect smaller perceptual incongruencies, thereby impacting the results of subjective measures.

These findings are consistent with the predictions of the CaP model [25], which suggests that perceptual incongruencies in VST AR have a stronger effect compared to the congruent VR condition; however, we found this only in objective measures. The CaP model currently addresses subjective higher-layer constructs, such as spatial presence. Hence, it focuses on consciously assessed effects (e.g., by questionnaire) and does not account for lower-layer effects, such as performance measures, which operate without cognitive intervention during both perception *and* evaluation. To provide a more comprehensive explanation of the observed effects, we propose to expand the CaP model to incorporate these lower-layer effects. By including mechanisms that account for performance measures and automatic unconscious processing, the model could better represent the full range of perceptual and cognitive phenomena not only in VST AR but in the whole XR spectrum.

6.5 Limitations and Future Work

Our experiment was conducted in the near-field distance, and the results are not necessarily transferable to the medium- or far-field distance as users rely on different depth cues depending on the distance [28]. Furthermore, we did not investigate the effect of the raw passthrough. This would have meant to include physical target objects to which participants would need to reach. Due to many potential confounds, including complexity when placing the object in relative proportion to the individual maximum arm-reach, and possible unwanted haptic feedback, we decided to use a combined approach that shows real-world reference objects through the passthrough (affected by ICD) and a virtual target object as overlay (affected by IPD). Hence, we covered a current relevant use case of AR by applying a composition of virtual and physical content. A future approach could implement an automated procedure to place physical target objects at certain distances.

Another limitation is the static ICD of $\approx$ 63 mm, allowing us to test only one fixed ICD with varying IPDs. Future work could explore adjustable ICDs using a custom HMD. However, our approach simulates a highly relevant real-world scenario, as HMDs with non-standard ICDs are atypical.

The generalizability of results for other HMDs is another limitation that needs to be addressed. We used the Meta Quest 3, which includes quite accurate tracking, high resolution, and a minimal amount of warping. It is not known which algorithms are applied to reduce the warping and to counteract possible lens distortions. Switching to a different HMD might change the results completely because of different hardware specifications and post-processing.

Confounds between VR and VST AR may have arisen due to different levels of embodiment, i.e., visibility of the own body. As Ries et al. [40] found out, embodying a virtual avatar improves distance estimations compared to not having a body. In our VST AR condition, participants were able to see their own bodies, while in VR, they did not, and only controller visibility was provided (see Fig. 1a). However, we believe that this effect was minimal, as participants had to do *blind* reaches, i.e., in the forward movement of the controller, the screen was blackened, and only in between were they able to inspect their embodiment. We also asked participants to keep their heads as still as possible. Thus, we assume that this visibility of the own body did not play a role in the task.

The study was conducted in two different countries. Even though we ensured the same strict experimental automated procedure, confounds might have occurred. Still, we would encourage this procedure as it fosters higher diversity and replicability in the results.

7 Conclusion

New use cases for VST AR arise due to an increase in the quality of the passthrough and hybrid usage with VR. However, VST AR technology is in comparison to VR or OST AR, underresearched. We need to consider perceptual incongruencies stemming from hardware limitations, alignment issues, etc., to provide a conclusive assessment of VST AR.

In this work, we examined incongruent stereoscopy, which arises due to user-individual IPD and fixed ICD in VST AR HMDs. We conducted a 2×3 mixed design empirical evaluation examining the effect in VR and VST AR as well as in three different phases, providing feedback or not. Results showed an increased underestimation in VST AR compared to VR as well as an effect of IPD-ICD mismatch. There were also significant interactions between the mismatch and the XR mode on both signed residual and proportional error. These findings provide evidence that incongruence of IPD and ICD in VST AR HMDs can influence near-field depth perception differently than in VR.

Notably, we found this evidence in objective measures but not in subjective measures, disclosing insights into human processing and deriving desiderata for integration and separation of high- and low-level effects in theoretical groundwork.

Overall, we believe that our findings are particularly important for the workplace usage of VST AR and specifically for serious use cases in medicine, industry, military, transportation, and other critical areas that rely on proper depth perception and near-field manipulation of objects. However, more research is needed to fully disclose the severity of IPD-ICD mismatch and to explore possibilities to counteract this issue. We encourage researchers to keep track of participants' IPD as it affects not only depth perception but potentially also overall performance and UX in VST AR and XR.

Acknowledgments

This research has been funded in part by the Bavarian State Ministry For Digital Affairs in the project XR Hub (Grant no. A5-3822-2-16) and the US National Science Foundation (CISE: IIS: HCC, Grant no. 2007435).

REFERENCES

[1] H. Adams, J. Stefanucci, S. Creem-Regehr, and B. Bodenheimer. Depth perception in Augmented Reality: The effects of display, shadow, and position. In *IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 792–801. IEEE, Piscataway, NJ, USA, 2022. doi: 10.1109/VR51125.2022.00101 2

[2] B. M. Altenhoff, P. E. Napieralski, L. O. Long, J. W. Bertrand, C. C. Pagano, S. V. Babu, and T. A. Davis. Effects of calibration to visual and haptic feedback on near-field depth perception in an immersive virtual environment. In *Proceedings of the ACM Symposium on Applied Perception*, pp. 71–78. ACM, New York, NY, USA, 2012. doi: 10.1145/2338676.2338691 2, 3, 8

[3] R. T. Azuma. A survey of Augmented Reality. *Presence: Teleoperators and Virtual Environments*, 6(4):355–385, Aug. 1997. doi: 10.1162/pres.1997.6.4.355 1, 2, 3, 8

[4] D. Bates, M. Mächler, B. Bolker, and S. Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, Oct. 2015. doi: 10.18637/jss.v067.i01 4

[5] G. P. Bingham and C. C. Pagano. The necessity of a perception–action approach to definite distance perception: Monocular distance perception to guide reaching. *Journal of Experimental Psychology: Human Perception and Performance*, 24(1):145, Nov. 1998. doi: 10.1037/0096-1523.24.1.145 2, 8

[6] S. Chakraborty, H. Finney, H. Gagnon, S. Creem-Regehr, J. Stefanucci, and B. Bodenheimer. Inter-pupillary distance mismatch does not affect distance perception in action space. In *ACM Symposium on Applied Perception 2024*, SAP '24, pp. 1–9. ACM, New York, NY, USA, 2024. doi: 10.1145/3675231.3675242 3

[7] J. E. Cutting and P. M. Vishton. Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In W. Epstein and S. Rogers, eds., *Perception of Space and Motion*, Handbook of Perception and Cognition, pp. 69–117. Academic Press (Elsevier), San Diego, CA, USA, 1995. doi: 10.1016/B978-012240530-3/50005-5 2

[8] B. Day, E. Ebrahimi, L. S. Hartman, C. C. Pagano, A. C. Robb, and S. V. Babu. Examining the effects of altered avatars on perception-action in Virtual Reality. *Journal of Experimental Psychology: Applied*, 25(1):1, Mar. 2019. doi: 10.1037/xap0000192 3

[9] N. A. Dodgson. Variation and extrema of human interpupillary distance. In *Stereoscopic Displays and Virtual Reality Systems XI*, vol. 5291, pp. 36–46. SPIE, Bellingham, WA, USA, 2004. doi: 10.1117/12.529999 3

[10] D. Drascic and P. Milgram. Perceptual issues in Augmented Reality. In *Stereoscopic Displays and Virtual Reality Systems III*, vol. 2653, pp. 123–134. SPIE, Bellingham, WA, USA, 1996. doi: 10.1117/12.237425 1, 2, 3, 8

[11] E. Ebrahimi, B. Altenhoff, L. Hartman, J. A. Jones, S. V. Babu, C. C. Pagano, and T. A. Davis. Effects of visual and proprioceptive information in visuo-motor calibration during a closed-loop physical reach task in immersive virtual environments. In *Proceedings of the ACM Symposium on Applied Perception*, pp. 103–110. ACM, New York, NY, USA, 2014. doi: 10.1145/2628257.2628268 3

[12] E. Ebrahimi, B. M. Altenhoff, C. C. Pagano, and S. V. Babu. Carryover effects of calibration to visual and proprioceptive information on near field distance judgments in 3D user interaction. In *2015 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 97–104. IEEE, Piscataway, NJ, USA, 2015. doi: 10.1109/3DUI.2015.7131732 3

[13] E. Ebrahimi, S. V. Babu, C. C. Pagano, and S. Jörg. An empirical evaluation of visuo-haptic feedback on physical reaching behaviors during 3D interaction in real and immersive virtual environments. *ACM Transactions on Applied Perception (TAP)*, 13(4):1–21, July 2016. doi: 10.1145/2947617 3

[14] F. Faul, E. Erdfelder, A.-G. Lang, and A. Buchner. G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2):175–191, May 2007. doi: 10.3758/BF03193146 3

[15] H. C. Gagnon, H. Finney, J. K. Stefanucci, B. Bodenheimer, and S. H. Creem-Regehr. Reaching between worlds: Calibration and transfer of perceived affordances from virtual to real environments. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 1011–1021. IEEE, Piscataway, NJ, USA, 2024. doi: 10.1109/VR58804.2024.00120 2, 3

[16] S. G. Hart and L. E. Staveland. Development of NASA-TLX (task load index): Results of empirical and theoretical research. In *Advances in Psychology*, vol. 52, pp. 139–183. Elsevier, San Diego, CA, USA, 1988. doi: 10.1016/S0166-4115(08)62386-9 5

[17] T. Hartmann, W. Wirth, H. Schramm, C. Klimmt, P. Vorderer, A. Gysbers, S. Böcking, N. Ravaja, J. Laarni, T. Saari, F. Gouveia, and A. Maria Sacau. The spatial presence experience scale (SPES): A short self-report measure for diverse media settings. *Journal of Media Psychology*, 28(1):1–15, Jan. 2016. doi: 10.1027/1864-1105/a000137 5

[18] S. Ishihara. *Tests for Colour-Blindness*. Handaya, Tokyo, Hongo Harukicho, Tokyo, 1917. 4

[19] J. A. Jones, J. E. Swan, G. Singh, E. Kolstad, and S. R. Ellis. The effects of Virtual Reality, Augmented Reality, and motion parallax on egocentric depth perception. In *Proceedings of the 5th symposium on Applied perception in graphics and visualization*, APGV '08, pp. 9–14. ACM, New York, NY, USA, 2008. doi: 10.1145/1394281.1394283 2

[20] J. W. Kelly. Distance perception in Virtual Reality: A meta-analysis of the effect of head-mounted display characteristics. *IEEE Transactions on Visualization and Computer Graphics*, 29(12):4978–4989, Dec. 2023. doi: 10.1109/TVCG.2022.3196606 2

[21] F. Kern and M. E. Latoschik. Reality stack i/o: A versatile and modular framework for simplifying and unifying XR applications and research. In *2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 74–76. IEEE, Piscataway, NJ, USA, 2023. doi: 10.1109/ISMAR-Adjunct60411.2023.00023 4

[22] H. K. Kim, J. Park, Y. Choi, and M. Choe. Virtual reality sickness questionnaire (VRSQ): Motion sickness measurement index in a Virtual Reality environment. *Applied Ergonomics*, 69:66–73, May 2018. doi: 10.1016/j.apergo.2017.12.016 5

[23] K. Kohm, S. V. Babu, C. Pagano, and A. Robb. Objects may be farther than they appear: Depth compression diminishes over time with repeated calibration in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics*, 28(11):3907–3916, Nov. 2022. doi: 10.1109/TVCG.2022.3203112 2, 8

[24] E. Kruijff, J. E. Swan, and S. Feiner. Perceptual issues in Augmented Reality revisited. In *IEEE International Symposium on Mixed and Augmented Reality*, pp. 3–12. IEEE, Piscataway, NJ, USA, 2010. doi: 10.1109/ISMAR.2010.5643530 1, 2, 3, 8

[25] M. E. Latoschik and C. Wienrich. Congruence and plausibility, not presence: Pivotal conditions for XR experiences and effects, a novel approach. *Frontiers in Virtual Reality*, 3, June 2022. doi: 10.3389/frvir.2022.694433 2, 3, 9

[26] W.-Y. Lin, Y.-C. Wang, D.-R. Wu, R. Venkatakrishnan, R. Venkatakrishnan, E. Ebrahimi, C. Pagano, S. V. Babu, and W.-C. Lin. Empirical evaluation of calibration and long-term carryover effects of reverberation on egocentric auditory depth perception in VR. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 232–240. IEEE, Piscataway, NJ, USA, 2022. doi: 10.1109/VR51125.2022.00042 2, 8

[27] D. Lüdecke, M. S. Ben-Shachar, I. Patil, P. Waggoner, and D. Makowski. performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, 6(60), Apr. 2021. doi: 10.21105/joss.03139 4, 5

[28] M. Mansour, P. Davidson, O. Stepanov, and R. Piché. Relative importance of binocular disparity and motion parallax for depth estimation: A computer vision approach. *Remote Sensing*, 11:1990, Aug. 2019. doi: 10.3390/rs11171990 2, 3, 9

[29] M. L. Mazow, T. C. Prager, and G. Cathey. Assessment of three stereo acuity tests. *American Orthoptic Journal*, 33(1):111–115, Apr. 1983. doi: 10.1080/0065955X.1983.11981608 4

[30] L. Meteyard and R. A. Davies. Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language*, 112:104092, Jan. 2020. doi: 10.1016/j.jml.2020.104092 5

[31] P. Milgram and F. Kishino. A taxonomy of Mixed Reality visual displays. *IEICE Transactions on Information Systems*, E77-D(12):1321–1329, Dec. 1994. 2

[32] M. Minsky. Telepresence. *OMNI magazine*, OMNI magazine:45–52, 1980. 2

[33] S. Nakagawa, P. C. Johnson, and H. Schielzeth. The coefficient of determination R 2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface*, 14(134):20170213, Sept. 2017. doi: 10.1098/rsif.2017.0213 5

[34] P. E. Napieralski, B. M. Altenhoff, J. W. Bertrand, L. O. Long, S. V. Babu, C. C. Pagano, J. Kern, and T. A. Davis. Near-field distance perception in real and virtual environments using both verbal and action responses. *ACM Trans. Appl. Percept.*, 8(3):18:1–18:19, Aug. 2011. doi: 10.1145/2010325.2010328 2

[35] C. C. Pagano and G. P. Bingham. Comparing measures of monocular distance perception: Verbal and reaching errors are not correlated. *Journal of Experimental Psychology: Human Perception and Performance*, 24(4):1037–1051, May 1998. doi: 10.1037/0096-1523.24.4.1037 2

[36] C. C. Pagano and R. W. Isenhower. Expectation affects verbal judgments but not reaches to visually perceived egocentric distances. *Psychonomic Bulletin & Review*, 15(2):437–442, Apr. 2008. doi: 10.3758/PBR.15.2.437 4

[37] K. Pfeil, S. Masnadi, J. Belga, J.-V. T. Sera-Josef, and J. LaViola. Distance perception with a video see-through head-mounted display. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI '21, pp. 1–9. ACM, New York, NY, USA, 2021. doi: 10.1145/3411764.3445223 2

[38] J. Ping, Y. Liu, and D. Weng. Comparison in depth perception between Virtual Reality and Augmented Reality systems. In *IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 1124–1125. IEEE, Piscataway, NJ, USA, 2019. doi: 10.1109/VR.2019.8798174 2

[39] R. S. Renner, B. M. Velichkovsky, and J. R. Helmert. The perception of egocentric distances in virtual environments - A review. *ACM Computing Surveys*, 46(2):23:1–23:40, Dec. 2013. doi: 10.1145/2543581.2543590 2, 3

[40] B. Ries, V. Interrante, M. Kaeding, and L. Anderson. The effect of self-embodiment on distance perception in immersive virtual environments. In *Proceedings of the 2008 ACM Symposium on Virtual Reality Software and Technology*, VRST '08, pp. 167–170. ACM, New York, NY, USA, 2008. doi: 10.1145/1450579.1450614 9

[41] J. J. Rieser. Development of perceptual-motor control while walking without vision: The calibration of perception and action. In *Sensory-Motor Organizations and Development in Infancy and Early Childhood: Proceedings of the NATO Advanced Research Workshop on Sensory-Motor Organizations and Development in Infancy and Early Childhood Chateu de Rosey, France*, pp. 379–408. Springer, Berlin, Germany, 1990. doi: 10.1007/978-94-009-2071-2_30 2, 3

[42] J. P. Rolland and H. Fuchs. Optical versus video see-through head-mounted displays in medical visualization. *Presence*, 9(3):287–309, June 2000. doi: 10.1162/105474600566808 2, 8

[43] D. Roth and M. E. Latoschik. Construction of the virtual embodiment questionnaire (VEQ). *IEEE Transactions on Visualization and Computer Graphics*, 26(12):3546–3556, Dec. 2020. doi: 10.1109/TVCG.2020.3023603 5

[44] M. Santoso and J. Bailenson. How video passthrough headsets influence perception of self and others. *Cyberpsychology, Behavior, and Social Networking*, Oct. 2024. doi: 10.1089/cyber.2024.0398 2

[45] T. Schubert, F. Friedmann, and H. Regenbrecht. The experience of presence: Factor analytic insights. *Presence: Teleoperators and Virtual Environments*, 10(3):266–281, June 2001. doi: 10.1162/105474601300343603 5

[46] G. Singh, J. E. Swan, J. A. Jones, and S. R. Ellis. Depth judgment tasks and environments in near-field Augmented Reality. In *2011 IEEE Virtual Reality Conference*, pp. 241–242. IEEE, Piscataway, NJ, USA, 2011. doi: 10.1109/VR.2011.5759488 4

[47] K. Stanney, C. Fidopiastis, and L. Foster. Virtual Reality is sexist: But it does not have to be. *Frontiers in Robotics and AI*, 7:4, Jan. 2020. doi: 10.3389/frobt.2020.00004 3

[48] J. Swan, M. Livingston, H. Smallman, D. Brown, Y. Baillot, J. Gabbard, and D. Hix. A perceptual matching technique for depth judgments in optical, see-through Augmented Reality. In *IEEE Virtual Reality Conference (VR 2006)*, pp. 19–26. IEEE, Piscataway, NJ, USA, 2006. doi: 10.1109/VR.2006.13 2

[49] K. Vaziri, M. Bondy, A. Bui, and V. Interrante. Egocentric distance judgments in full-cue video-see-through VR conditions are no better than distance judgments to targets in a void. In *IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 1–9. IEEE, Piscataway, NJ, USA, 2021. doi: 10.1109/VR50410.2021.00056 2

[50] C. C. Voeten. Using 'buildmer' to automatically find & compare maximal (mixed) models, 2020. 4

[51] F. Westermeier, L. Brübach, M. E. Latoschik, and C. Wienrich. Exploring plausibility and presence in Mixed Reality experiences. *IEEE Transactions on Visualization and Computer Graphics*, 29(5):2680–2689, May 2023. doi: 10.1109/TVCG.2023.3247046 2

[52] F. Westermeier, L. Brübach, C. Wienrich, and M. E. Latoschik. Assessing depth perception in VR and video see-through AR: A comparison on distance judgment, performance, and preference. *IEEE Transactions on Visualization and Computer Graphics*, 30(5):2140–2150, May 2024. doi: 10.1109/TVCG.2024.3372061 2, 3, 4, 8

[53] H. Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2009. 4

[54] C. Wienrich, P. Komma, S. Vogt, and M. Latoschik. Spatial presence in Mixed Realities – Considerations about the concept, measures, design, and experiments. *Frontiers in Virtual Reality*, 2, Oct. 2021. doi: 10.3389/frvir.2021.694315 5, 9

[55] P. Willemsen, M. B. Colton, S. H. Creem-Regehr, and W. B. Thompson. The effects of head-mounted display mechanics on distance judgments in virtual environments. In *Proceedings of the 1st Symposium on Applied Perception in Graphics and Visualization*, APGV '04, pp. 35–38. ACM, New York, NY, USA, 2004. doi: 10.1145/1012551.1012558 2

[56] P. Willemsen, A. A. Gooch, W. B. Thompson, and S. H. Creem-Regehr. Effects of stereo viewing conditions on distance perception in virtual environments. *Presence: Teleoper. Virtual Environ.*, 17(1):91–101, Feb. 2008. doi: 10.1162/pres.17.1.91 3, 4

[57] B. H. Woldegiorgis, C. J. Lin, and W.-Z. Liang. Impact of parallax and interpupillary distance on size judgment performances of virtual objects in stereoscopic displays. *Ergonomics*, 62(1):76–87, Jan. 2019. doi: 10.1080/00140139.2018.1526328 3