The State of Replication at IEEE ISMAR and IEEE VR: A Scoping Literature Review (2010 – 2024) and Online Survey

Mohammed Safayet Arefin*
Colorado State University, USA

Verena Biener†
University of Stuttgart, Germany

S M Rashidul Hasan Nijhum‡
Colorado State University, USA

Florian Weidner§
University of Glasgow, UK

Jens Grubert¶
Coburg University, Germany

J. Edward Swan II‖
Mississippi State University, USA

Abstract

A replication attempts to confirm outcomes of earlier research, which is critical in validating and generalizing scientific findings. Yet, its prevalence and practices remain underexplored in Augmented and Virtual Reality (AR/VR) research. To address this, we present a scoping literature review of replication studies within IEEE ISMAR and IEEE VR, spanning 15 years from 2010 to 2024. Our analysis revealed that replication in AR/VR research is rare. Of 2167 total papers reviewed, less than 4% were identified as replication studies. Of these, conceptual replications were the predominant type (57%). Most of those were studies in VR (67%) and within-subject designs (66%). Complementing the literature survey results, we conducted an online survey with 61 participants about their experiences with replication studies and found that 39% of them had conducted a replication study. However, limited resources and external motivation hamper the execution of replication studies among these AR/VR researchers. Combining the findings from our literature review and online survey, we discuss the current state of replication research and the factors contributing to its infrequency. We provide recommendations to improve AR/VR research replication practices, focusing on research culture and reporting, and discuss ongoing challenges.

Index Terms: Replication, Survey, Virtual Reality, Augmented Reality, Extended Reality

1 Introduction

Empirical research largely falls into exploratory and confirmatory studies. Exploratory studies examine potentially new concepts in a research domain to develop new hypotheses. Confirmatory studies aim to test hypotheses, and a subset of those studies intends to verify existing findings by replicating earlier research, which is essential to the scientific process. They ensure reliable and generalizable results and explore alternative explanations [80].

In augmented reality (AR) and virtual reality (VR), systems are complex and exhibit strong dependencies on hardware. Moreover, over the past 15 years, AR and VR research has advanced significantly, driven by improvements in head-mounted displays, rendering techniques, robust calibration methods, and artificial intelligence. In this evolving landscape, confirming previous findings is essential to ensure that new work builds on valid and generalized findings. Despite the generally accepted significance of replication and the added challenges of the AR/VR domain, replication in AR/VR work appears to be underrepresented. This raises concerns about the scientific robustness and empirical foundation of the field.

Therefore, in this paper, we have two aims: First, we aim to assess the current state of replication research in AR/VR, investigating the number of replication studies published, the methods used in this research, replication types, confirmation, the researchers who conduct it, and other key findings. Second, we are interested in our communities' experiences and thoughts on replication studies. We aim to understand perspectives on replication research, the reasons behind replication studies, their perceived value, and the factors preventing them. Overall, our paper represents the first effort to explore the status and scope of replication studies in the AR/VR community, along with researchers' perspectives and experiences.

To this end, we performed a scoping literature survey of 15 years of publications from the IEEE International Symposium on Mixed and Augmented Reality (IEEE ISMAR) and IEEE International Conference on Virtual Reality and 3D User Interfaces (IEEE VR), ranging from 2010 to 2024. In total, we considered 2167 publications for our survey. 4% (83 out of 2,167) are replication studies. Furthermore, an online survey of 61 academics indicates a high perceived value for replication research (5.54 on a scale from 1 to 7, with 7 being very strong support). 39% of these academics reported having conducted at least one replication study. The survey respondents mentioned several primary obstacles to replication research, including limited resources (funding, time, and data) and academic challenges (perceived value, difficulties in publication, and insufficient encouragement). Our main contributions are:

1. A publicly available collection of replication studies in IEEE ISMAR and VR from 2010 to 2024, informing and guiding researchers when doing replications.
2. An analysis of all replications and insights from an online survey about perceived values and barriers.
3. Recommendations to promote replications by improving research culture, reporting, and community efforts.

Without replication studies, trust in the findings of previous work is limited and easily eroded. This work provides an in-depth examination of the research culture within IEEE ISMAR and VR, aiming to enhance awareness and offer insights to inform future studies. Our findings are valuable for institutions, journals, conferences, and funding agencies to adjust their policies. They can utilize our work to inform policy changes, encourage replication, require open data, and establish specific programs and platforms targeted at replication research. Through this work, we aim to contribute to AR/VR research by promoting higher research transparency, stronger validation of findings, and, ultimately, more reliable and efficient progress.

2 Related Surveys

Primary research fields closely connected to the AR/VR domain, such as human-computer interaction (HCI), human factors, and psychology, have already explored the state of replication studies.

*e-mail: arefin@acm.org
†e-mail: verena.biener@visus.uni-stuttgart.de
‡e-mail: smrashidulhasannijhum@gmail.com
§e-mail: florian.weidner@glasgow.ac.uk
¶e-mail: jens.grubert@gmail.com
‖e-mail: swan@acm.org

First two authors contributed equally to this work.

Human-computer Interaction (HCI): HCI has shown awareness of replications through publications [44, 37], workshops [112], and panels [113]. To our knowledge, only Hornbaek et al. [44] conducted a replication survey in HCI. They analyzed publications from 2008 to 2010 in the ACM Conference on Human Factors in Computing Systems (CHI), ACM Transactions on Computer-Human Interaction (TOCHI), Human-Computer Interaction (HCI), and the International Journal of Human-Computer Studies (IJHCS). Results indicated that 3% of the papers engaged in HCI-based replication studies. Among these, approximately 50% successfully replicated the original work. The survey did not categorize HCI papers by topics, so it remains unclear if there are any AR/VR papers present in the HCI replication survey.

Human Factors: Within human factors research, Jones et al. [48] investigated replications by analyzing the 1991 issue of the *Journal of Human Factors*. Their study involved examining eight articles as parent articles, along with child articles that had cited these parent articles from 1991 to September 2006. They identified the presence of replication efforts, comprising 30 replication papers, within the human factors field. Moreover, 63% of the replication studies were performed by authors different from those of the original parent articles. Similar to the HCI replication survey, Jones et al. [48] also omitted to specify if any papers involved AR/VR-based replication.

Psychology: In 2012, Makel et al. [72] provided a psychology replication overview by analyzing the 100 psychology journals with the highest 5-year impact factors since 1900. Analyzing 500 random articles, they found that the rate of replications was 1%. They observed successful replication in 90% of the cases where authorship overlapped between the original and replication. The Reproducibility Project in the field of Psychology [19] is a notable publication that analyzed original versus replication studies focusing on p values, effect sizes, replication team assessments, and effect size meta-analysis. This project replicated 100 experimental and correlational studies from three psychology journals using materials from the original authors. Many replications showed weaker support for the initial findings: 97% of original studies were significant ($p < .05$), while 36% of replications were significant. Replication success was mainly associated with the strength of the original study's evidence, like the initial p-value, rather than the research team's expertise in the field. However, Patil et al. [85] noted that media often focus on the 36% replication success rate from the *Reproducibility Project: Psychology* [19]. They claim that this does not truly represent replication efforts and recommend using 95% prediction intervals instead of the traditional $p < 0.05$, which ignores sampling variability. This highlights an ongoing discussion in differing methodological recommendations on reporting the statistics. Some methodologists suggested shifting from the null-hypothesis significance testing (NHST) to the estimation of effect sizes, confidence intervals, and meta-analysis, while reporting the analysis of the study [21]. On the other hand, some methodologists recommend using NHST together with effect sizes and confidence intervals [59]. Within the CHI community, methodologists have made multiple attempts to move away from NHST [2, 24]. Besanon and Dragicevic [10] conducted a survey to determine the impact of these efforts on CHI researchers and found none. Wagenmakers et al. [108] proposed preregistering studies with their analysis plans to enhance research transparency, potentially improving replication success rates in psychology. Moreover, to reduce the file drawer problem (A bias toward publishing significant results), Simonsohn et al. [98] proposed p-curve values. Interestingly, Nosek and Lakens [81] noted that psychology journal editors and reviewers frequently dissuaded the publication of replication studies. To promote replication publications in Psychology, they initiated the first known journal issue in the Journal of *Social Psychology* titled *Registered Reports: A Method to Increase the Credibility of Published Results*. This special issue included 15 articles replicating significant findings in social psychology. Yet, none of the psychology replication publications specifically concentrated on AR/VR research.

AR/VR: AR/VR researchers have conducted numerous user-study-oriented surveys [101, 22]. However, to our knowledge, currently, no work addresses the scope of replication in AR/VR. Still, AR/VR researchers have begun emphasizing the importance of community replication through special paper submission tracks, workshops, and tutorials. The Workshop on Replication in Extended Reality (WoRXR) was held four times at IEEE ISMAR from 2021 to 2024. Twelve replication papers have been published in this workshop altogether. Since 2018, several tutorials have been presented multiple times at IEEE VR and IEEE ISMAR. The ACM Symposium on Virtual Reality Software and Technology (VRST) recently introduced a "Reproduction Challenge" submission track in 2024 [1]. Although the track supports both *replication papers* and *replication posters*, only four posters were initially published, indicating a lack of awareness about the significance of replication research in the AR/VR community. The track reopened for submissions in 2025 [2]. The Virtual Experience Research Accelerator (VERA) [3] project has launched, which aims to provide tools to AR/VR researchers to improve the quality of user-based research and replication. The full impact of VERA has not been determined yet.

To our knowledge, currently, there are no surveys available that address the scope of replication in the domains of AR/VR research. Therefore, the frequency of replication over time, replication types, topic confirmation, researchers, environment, design, display, and community insights in the AR/VR research domain is still unknown. Still, based on the overview of replication studies and community insights, actionable recommendations to foster replication work are lacking, which limits trust in prior findings.

3 Survey Methodology

Following the PRISMA protocol [104], our scoping literature review systematically explores the extent of replication in user-based studies within IEEE ISMAR and VR. It aims to classify these and identify the AR/VR topics, researchers, environment, design, participants, and display of replication studies. The PRISMA checklist is available in the supplementary material.

3.1 Identification

We identified replication studies by analyzing publications from key venues, a method previously used in fields like HCI [44] and psychology [77]. We focused on full papers published at IEEE ISMAR and IEEE VR, excluding pre-2018 IEEE 3DUI papers. Although other platforms like ACM CHI and ACM VRST also publish AR/VR research, we selected IEEE VR and IEEE ISMAR as the two premier venues (both ranked as A* in CORE 2023) that focus explicitly on AR/VR research and represent a central part of the community. Future replication research should expand the analysis to include more venues, including CHI, VRST, but also others such as TOCHI, Frontiers, MIT Presence, and IJHCI.

Our analysis covered the last 15 years (2010–2024), aligning with major technological advances in AR/VR (among others, hardware, software, interaction techniques, calibration procedures, and tracking algorithms). With this scope in mind, we manually downloaded all articles from the IEEE Digital Library (N = 2,167, cf. Figure 1, step "Identification"). We did not use the search functions of the IEEExplore. Instead, we manually accessed the con-

[1] https://vrst.hosting.acm.org/vrst2024/reproduction-challenge/, accessed 2025-06-16

[2] https://vrst.acm.org/vrst2025/index.php/reproduction-challenge/, accessed 2025-06-16

[3] https://sreal.ucf.edu/vera/, accessed 2025-04-11

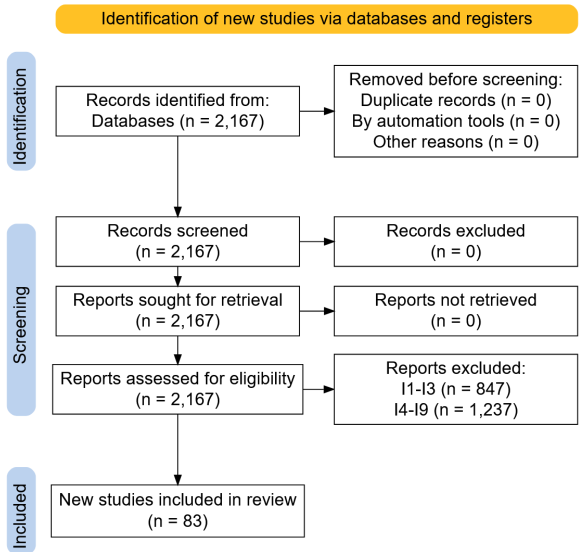

Figure 1: PRISMA diagram.

ference and journal proceedings. All papers accessible online on 14.01.2025 are included in the review.

3.2 Screening: Checking Eligibility

After retrieving all 2.167 papers, we checked them for eligibility. The following sections explain the two-step process in detail.

3.2.1 Step 1: Context

Following the eligible criteria for replications in HCI, by Hornbaek et al. [44], we defined the following criteria for a paper to be considered eligible within the AR/VR domain before evaluating its replication potential.

I1) Inclusion of AR/VR Technology: The paper must use AR/VR technologies. We excluded AR/VR papers that did not involve hands-on experience using AR/VR technology, such as those conducted in online surveys.
I2) User-Based Experiment: A paper must feature a user study. While system and model papers with purely technical evaluations constitute significant AR/VR research, we focus on replication studies involving human subjects, as human interaction introduces significant variance and is a crucial factor in understanding AR/VR experiences. These works benefit greatly from validation through replication. Other disciplines have also applied this approach [44, 77].
I3) Inclusion of Quantitative Data: The paper must report quantitative data and report on a statistical analysis. We excluded purely qualitative studies as their interpretation is subjective and, thus, harder to compare systematically.

Based on these criteria, which we considered combined in a first step, we discarded 847 papers.

3.2.2 Step 2: Replication Studies

We then checked the remaining 1320 papers to see if they were replication studies. Here, we applied an additional set of inclusion criteria: For each paper, we verified whether

I4) it cites the original paper it aims to replicate and
I5) it seeks to confirm, validate, expand, or generalize the original study's findings, as defined by Hornbaek et al. [44].

These criteria aim to establish a direct link between an original piece of work and the current work. In addition, we verified that

I6) it provides evidence of methodological consistency with previous work or
I7) includes deliberate variations designed to test the robustness of results and
I8) does not coincidentally produce similar findings without explicit intent to replicate and
I9) is not an intra-experiment replication. If a study includes two experiments where the second is described as replicating the first with modifications, we do not classify it as a replication in our survey. It lacks comparison with external experiments. A similar approach was used by Hornbaek et al. [44] in HCI.

These criteria verify that all papers in our corpus are replication studies and that authors do not mistakenly claim to replicate prior work. We apply these criteria by first searching for keywords within the paper's abstract, introduction, experimental task, setup, design, and discussion, and then manually checking these sections to agree with the criteria I4–I9. Generally, a replication paper is anticipated to include the term *Replication*. However, researchers reported that many replication studies do not contain this term [72, 77, 44]. Therefore, we also considered further search terms: 'reproduce', 'follow-up study', 'earlier experiment', 'similar to', 'previous experiment', 'reproduce results', 'inspired by', 'adapted from'. In this step, considering all criteria, we excluded 1.237 papers. The final corpus consists of 83 replication studies (cf. Figure 1, "Included").

3.3 Data Extraction

After identifying the replication studies, we categorize the replication papers using a replication typology. Based on our knowledge, there are no universally accepted typologies for replication. In this paper, we focus on Hendrick's classification [42] as it is most frequently referenced [35]:

- Strict Replication: This involves replicating the original study with identical methods, procedures, measures, and context, ensuring that both studies have the same independent and dependent variables.
- Partial Replication: This involves replicating a study with deliberate or minor modifications to the original to examine other influencing factors.
- Conceptual Replication: This type uses a different technique or method to determine whether similar results can be achieved using an alternative experimental approach.

In addition, we extracted other information from each paper, including the environment (such as AR and VR), topic (e.g., perception, interaction, and others), details on the experimental design, demographic information (such as the participants' count, gender, and age), the technical setup employed (e.g., display used), and whether the same or different researchers replicated the original study. There was no critical appraisal of the replications' quality.

3.4 Survey Process

We divided the screening task among four authors. Each author reviewed a subset of papers to check for eligibility. Before beginning the review, each author exchanged at least one coded paper with other authors to verify reliability. During regular meetings, the authors discussed articles that warranted additional discussion. We stored information in a shared spreadsheet. The spreadsheet is attached as supplemental material, containing detailed information on our survey.

4 Overview of Replication Survey

In this survey, we examined a collection of 2167 publications, including journal and conference papers from the IEEE ISMAR and

Table 1: Replication papers split by venue and type from 2010–2024.

Venue	Type	Total Papers	Eligible	**Replications**
IEEE ISMAR	Journal (TVCG)	264	134 (51%)	9 (3%)
IEEE ISMAR	Conference	713	412 (58%)	23 (3%)
IEEE VR	Journal (TVCG)	401	270 (67%)	14 (3%)
IEEE VR	Conference	789	504 (64%)	37 (5%)
	Total	**2167**	**1320 (61%)**	**83 (4%)**

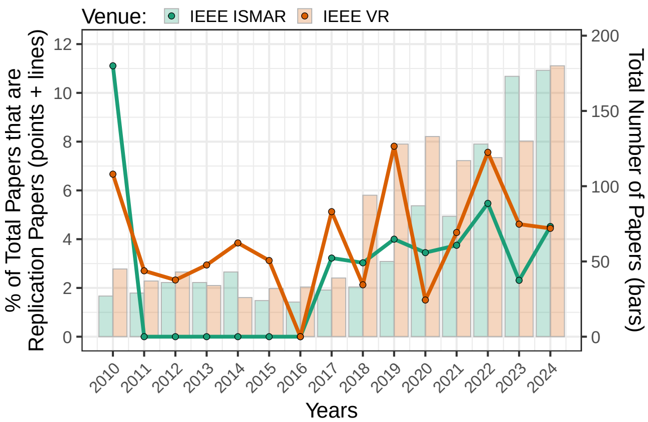

Figure 2: Replication papers per venue over time. The high percentages in 2010 are artefacts due to few publications. From 2016 onward, numbers have been slightly increasing.

IEEE VR conferences, spanning from 2010 to 2024. Of those, 83 (4%) are replication papers. In the following, we provide an overview of these 83 papers, looking at frequency, replication type, topics, whether the replication confirmed initial findings, whether it is in the AR or VR domain, the study design, participants, and what displays the paper used. When applicable, we report the replications' details and compare the replications' methodology with the original work. Please refer to the supplemental material for a comprehensive list of replication papers, their assigned categories, and the corresponding original papers.

4.1 Frequency of Replication Over Time

Table 1 summarizes the distribution of papers among outlets. In the past 15 years, ISMAR conferences have published 32 replication papers, which represent 3% of their total papers, while IEEE VR has published 51 replication papers, representing 4% of their total papers. Although IEEE VR has a higher number of replications, the proportion relative to the overall number of papers remains similar.

Figure 2 shows the frequency of replication papers as a percentage of the total papers from 2010 to 2024. The number of replication papers relative to the total number of papers ranges from 0% to 11% for ISMAR, 0% to 8% for IEEE VR, and 0% to 8% for ISMAR and IEEE VR combined. We do not observe consistent trends. However, there appears to be an increase in recent years.

An important observation is that, despite the annual increase in the total number of papers for both conferences, the proportion of replication papers relative to the total remains very low each year. For example, in 2010, the total number of papers was 27, with the replication papers comprising 11% for ISMAR, and for IEEE VR, the replication papers made up 7%, with a total of 45 papers. By 2024, the total number of papers rose to 177 for ISMAR (more than 6 times that published in 2010), with replication articles still comprising only 5%, and for IEEE VR, the total number of papers increased to 180 articles (4 times that published in 2010), replication articles comprising only 4%.

Figure 3 illustrates the frequency of replication papers as a percentage of the overall papers from 2010 to 2024 in journal and conference publications. Our analysis revealed nearly similar percentages of replication papers: 3% of 665 journal papers versus 4% of 1502 conference papers. No consistent patterns are evident in the frequency of replication papers compared to the total number of publications across journal and conference formats, although a slight increase seems to have happened in recent years. Overall, the number of replication papers is 4%, which is relatively low and barely differs between venues (ISMAR and VR) and publication types (journal vs. conference paper).

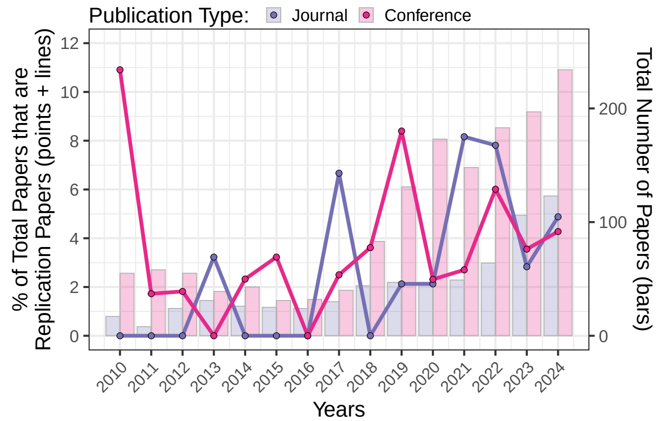

Figure 3: Percentage of replication papers by publication type (Journal or Conference) over time. The high percentage in 2010 is due to an overall low number of publications. From 2016 onward, numbers are slightly increasing.

On average, the time between the publication of the original and the replication study was 5.7 years ($sd = 6.8$). 80% of replications were conducted within 10 years after the original study, and another 19% within 25 years. The longest gap between the original paper and the replication was 45 years for one of the papers.

4.2 Replication Types

In our survey, out of 83 replication papers, 5 papers (6%) are categorized as strict replications, 31 papers (37%) are partial replications, and the majority, 47 papers (57%), are considered to be conceptual replications (see Table 2).

An instance of *strict replication* is the study on locomotion interfaces by Kelly et al. [50]. The primary authors repeated their original research [51] where they examined the effects of different teleportation interfaces on users' capacity to navigate and orient in VR within a laboratory setting. The objective was to determine whether the original study's outcomes remained consistent when participants employed their personal VR devices in remote environments instead of controlled laboratory settings. They employed an identical triangle completion task to evaluate spatial updating during teleportation for participants at remote locations. They assessed similar errors, including absolute distance error, angular error, axial error, and response latency — independent variables remained consistent. The replication paper also compared its findings with the initial laboratory data utilizing comparable statistical methods.

An instance of *partial replication* is illustrated by the research on AR display context switching and focal display switching carried out by Arefin et al. [3]. They replicated research by Gabbard et al. [28], which explored the effects of AR display context switching and focal distance switching on human performance utilizing a text-based visual search task with a Micro Vision Nomad AR display. The replication study [3] pursued identical research objectives, employing the same experimental task and methodology but involving additional experimental variables and a different AR display, termed the AR Haploscope. Although the methodologies of the original and replication studies are alike, the experimental displays

Table 2: All papers split by replication type.

Type	Paper
Strict (N=5)	Kelly et al. [50], Venkatakrishnan et al. [105], Wang et al. [109], Ye et al. [115], Yin et al. [116]
Partial (N=31)	Arefin et al. [3], Batmaz et al. [4], Batmaz et al. [6], Batmaz et al. [6], Baumeister et al. [8], Casallas et al. [17], Cortes et al. [20], Do et al. [23], Gagnon et al. [30], Gagnon et al. [29], Grndling et al. [38], Hamada et al. [40], Jeong et al. [47], Jung et al. [49], Laha et al. [57], Lee et al. [61], Li et al. [64], Li et al. [63], Liu et al. [68], Mahmud et al. [71], Marquardt et al. [73], Marquardt et al. [74], Otono et al. [84], Prytz et al. [90], Raikwar et al. [91], Rietzler et al. [93], Todd et al. [102], Volmer et al. [107], Wolf et al. [114]
Conceptual (N=47)	Ahmed et al. [1], Arefin et al. [3], Batmaz et al. [4], Batmaz et al. [6], Batmaz et al. [6], Benda et al. [9], Bhargava et al. [11], Blum et al. [12], Bodenheimer et al. [13], Buck et al. [14], Casallas et al. [17], Chiossi et al. [18], Cortes et al. [20], Do et al. [23], Eom et al. [25], Faleel et al. [26], Fukushima et al. [27], Gagnon et al. [30], Gagnon et al. [29], Gandy et al. [32], Gao et al. [33], Garrido et al. [34], Gower et al. [36], Grndling et al. [38], Hamada et al. [40], Hayashi et al. [41], Hoinville et al. [43], Ibrahim et al. [45], Jarrell et al. [46], Jeong et al. [47], Jung et al. [49], Khan et al. [52], Klose et al. [53], Kodama et al. [54], Krogmeier et al. [56], Laha et al. [57], Lee et al. [61], Lenz et al. [62], Li et al. [64], Li et al. [63], Lin et al. [65], Lin et al. [66], Liu et al. [68], Liu et al. [67], Lougiakis et al. [69], MacQuarrie et al. [70], Mahmud et al. [71], Marquardt et al. [73], Marquardt et al. [74], Moraes et al. [75], Nilsson et al. [78], Ogawa et al. [83], Ogawa et al. [82], Otono et al. [84], Peck et al. [86], Peer et al. [87], Perrin et al. [88], Phillips et al. [89], Prytz et al. [90], Raikwar et al. [91], Richard et al. [92], Rietzler et al. [93], Rosales et al. [94], Shen et al. [96], Sigurdarson et al. [97], Souchet et al. [99], Todd et al. [102], Torres et al. [103], Volmer et al. [107], Weidner et al. [110], Wienrich et al. [111], Wolf et al. [114], You et al. [118], Yu et al. [119], Yoon et al. [117], Zenner et al. [120]

Note: Due to space constraints, author names are only cited here.

and some experimental variables differ; thus, based on replication criteria and their properties, this constitutes a partial replication.

An example of *conceptual replication* is the study conducted by Klose et al. [53], which investigates the impact of various text representations in AR within dual-task scenarios. It replicates the research initially performed by Rzayev et al. [95]. Both original and replication address the same research question. The original analyzed three different styles of text representation in AR for conditions that involve walking and sitting. Klose et al. [53] conducted their replication with two alternative text representation styles (distinct from the original) in walking scenarios. Consequently, both studies focus on the same research question but adopt divergent methodologies and experimental frameworks and introduce novel variables to explore the question from an alternative angle.

4.3 AR/VR Topic of Replication

We inductively classified all 83 replications and identified 8 topic areas. Two steps were used to identify the topics: (1) An author initially proposed the topic areas. (2) Through group discussions, the authors finalized eight topic areas. The final topics are cognition and perception, locomotion, collaboration, interaction, interface design, human factors, simulation and modelling, and avatars (see Table 1 in Appendix A). Among the 8 topics, cognition and perception have the highest number of replication papers, 24 (29%) of the total replication papers. Collaboration has the lowest number of replication papers, with 3 (4%) of the total replication papers.

An example of replication within the *cognition and perception* domain (24 or 29% of the total replication papers) is the causal perception study conducted by Wang et al. [109]. They replicated the research conducted by Guski and Troje [39] to confirm the experimental results by extending the experiment to the VR environment.

In the field of *human factors* (14 or 17% of the total replication papers), a replication study example is the work by Moraes et al. [75]. The researchers explored the impact of auditory and audiovisual distractions on task performance within a VR-based auditory selective attention task, replicating the experimental methodology previously developed by their research group (Moraes et al. [76]).

An instance of a replication study in the *simulation and modeling* field (11 or 13% of the total replication papers) is the VR-based simulation for visual analytics research executed by Laha et al. [57]. In this study, researchers replicated their previous experiment, which utilized a CAVE-based VR system for volume data simulation [58]. In the replication study, the authors used a VR system based on a head-mounted display (HMD) for the simulation to examine the impact of immersion on visual analysis tasks.

An instance of a replication study within the *interaction* domain (10 or 12% of the total replication papers) is the work by Casallas et al. [17]. During their examination of interaction techniques to select moving targets, Casallas et al. [17] replicated the methodologies of their previous study [16], incorporating additional variables and features to improve the model for the selection task.

In the domain of *avatars* (9 or 11% of the total replication papers), Yu et al. [119] utilized a conceptual replication method to examine the variances in using a point cloud reconstruction-based avatar versus a virtual character-based avatar for representing a collaborator. This study conceptually replicated the work by Gamelin et al. [31], who initially contrasted point cloud representations obtained from a single depth camera with avatars that were 3D reconstructed without including facial or finger animations. The replication study incorporated real-time environment capture and live user capture in a point cloud, as well as facial expressions and finger tracking for avatar animation.

In the domain of *locomotion* (6 or 7% of the total replication papers), an example is the investigation of redirected walking conducted by Rietzler et al. [93]. During their study on curvature and bending gains in redirected walking, this publication partially replicated previous research by Langbehn et al. [60].

Within the field of *interface design* (6 or 7% of the total replication papers), Batmaz et al. [7] conducted a replication study that explores various grip styles. The researchers partially reproduced and expanded on previous work by Batmaz et al. [5] to assess the impact of different grip styles and examine how vergence-accommodation conflict (VAC) affects user performance.

An illustration of replication for *collaboration* (3 or 4% of the total replication papers) can be found in the study by Prytz et al. [90], where researchers explored the impacts of eye contact patterns in a collaborative task utilizing AR, replicating the methodology outlined by Nilsson et al. [79].

4.4 Replication Confirmation

Nosek et al. [19] replicated 100 studies from leading psychology journals. In the language of this paper, these could be considered *strict replications*. Each replicated study resulted in a single hypothesis test and a single p value. By definition, a successful study met the criterion of $p \leq .05$, indicating statistical significance. A central finding is that while 97% of the original studies had statistically significant results, only 36% of the replications had statistically significant results [19]. This low degree of reproducibility was attributed to biases in the original papers, where results with $p \leq .05$ were preferentially selected for analysis, reported in the final paper, and subsequently published. In contrast, for the replications, all results were analyzed, reported, and published.

Inspired by this earlier work, we decided to perform a similar analysis with the 5 strict replication papers. Among these replication papers [109, 115, 50, 116, 105], four reported a series of hypothesis tests that generated p values that could be directly compared between the replication paper and the original paper. The fifth, Yin et al. [116], fitted a model where success was not assessed

The final version of this record is available at: 10.1109/ISMAR67309.2025.00086

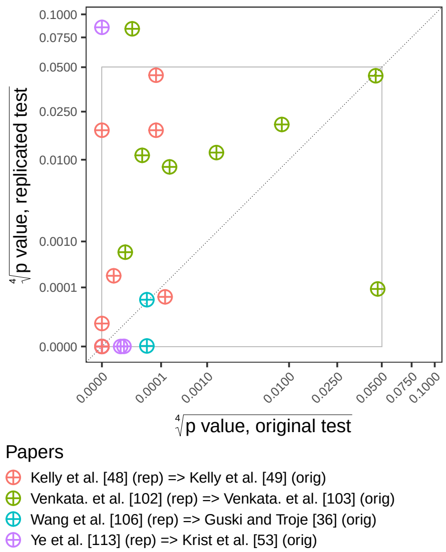

Figure 4: Comparison of p values for the strict replication papers: [50] (rep) ⇒ [51] (orig); [105] (rep) ⇒ [106] (orig); [109] (rep) ⇒ [39] (orig); [115] (rep) ⇒ [55] (orig). Because 13 points have very small values, the axes have a transformation of $\sqrt[4]{p}$. Yet, still five values are overlapping at the graph origin at $(0,0)$. The box encloses results where $p < .05$ for all papers. The dashed line indicates equivalent statistical evidence for both results.

with a hypothesis test.

Overall, between the four replication and original papers, we found 24 hypothesis tests that examined the same dependent variable against similar experimental conditions, and therefore could generate p values that could be directly compared. Instead of the p value category reported in these papers (typically levels such as $p < .01, p < .001$, etc.), we used code from the supplementary materials provided by OSC [19] to directly calculate p from the given hypothesis test (either an F test or a Chi-squared test). Fig. 4 shows the results, with the original p values along the x axis and the replicated p values along the y axis.

The value of p can be considered a measure of the strength of the statistical evidence (OSC [19]), with lower values indicating stronger evidence. The dashed line at 45° indicates equivalent statistical evidence for the replication and original result. Relative to the original result, points below this line indicate increased statistical evidence for the replication, while points above this line indicate weaker statistical evidence. Fig. 4 indicates that $5/24 = 21\%$ of the replications had stronger statistical evidence, $7/24 = 29\%$ had equivalent statistical evidence (including the 5 overlapping points at the origin), and $12/24 = 50\%$ had weaker statistical evidence, including a non-significant replication for two results.

Similar to the results reported by OSC [19], considering the 12 points above the 45° line, there is weaker statistical evidence for the replications. However, among the eight papers analyzed in Fig.4, for both the original and replication papers, the only hypothesis tests listed are those for which $p \leq .05$. The papers generally mention non-significant tests but do not give the test statistics, making it impossible to calculate the non-significant p values. Therefore, similar to OSC [19], the reported statistics appear to have similar selection biases among these papers.

4.5 Researchers of Replication Study

Of the 83 replication papers, 49 (59%) were replicating studies of different authors (different group), while 34 (41%) replicated prior studies of their group (see Table 3 in Appendix A). We considered a replication paper from the same group if it had at least one common author with the original paper. This means that studies by different authors are slightly more often replicated.

4.6 Environment of Replication Study

We categorized the environment of the replication study into AR, VR, and both environments. We found that 21 replication papers (25% of the total replication papers) were done in AR, 58 replication papers (70% of the total replication papers) in VR, and 4 replication papers (6% of the total replication papers) used both AR and VR (see Table 2 in Appendix A). When comparing whether the replication used the same environment as in the original study, we found that 47 replication papers (57% of the total replication papers) used the same environment, and 36 (43% of the total replication papers) used a different environment.

4.7 Design of Replication Study

Regarding the study design, we found that a within-subjects design was utilized in 55 replication studies (66%). Alternatively, 16 studies (19%) employed a between-subjects design, and 12 (14%) used a mixed design (see Table 4 in Appendix A). The percentage values closely resemble the trend found in non-replication AR usability studies [22]. When comparing the experimental design of the replication studies with the original studies, 45 replication papers (54%) maintained the same experimental design as the originals, while 38 replications (46%) opted for a different design.

4.8 Participants of Replication Study

42 replication studies (51% of all replication papers) included more participants than the original, 9 replication papers (11%) had an equal number of participants, and 26 replication studies (31%) included fewer participants. The remaining papers did not give enough details to compare participants. Additionally, the average sample size of AR/VR replication studies is 39 (within-subject: 36, between-subject: 59, mixed design: 53), whereas the average sample size of original studies is 35 (within-subject: 28, between-subject: 53, mixed design: 26). Based on the survey on CHI 2014 proceedings [15], in the HCI community, the average sample size of in-person experimental studies (non-replication) is 20 (within-subject: 17, between-subject: 26, mixed design: 25). Therefore, AR/VR replication studies typically consider a larger sample size compared to (non-replication) HCI experimental studies.

Considering the gender of the participants in the replication research, we observed that 74 articles included both male and female subjects (Male: 52% and Female: 48%). In comparison, the original studies had an average of 44% females and 56% males. In general, AR/VR replication studies involved an almost equal number of male and female participants, which contrasts with the male-dominated gender distribution (64% male, 36% female) in AR usability studies [22]. Of all replication papers, 66 (80%) provided data on the mean age of the participants. The weighted mean age of participants in the replication studies was 25.61 years. This suggests that most replication studies have focused on younger participants rather than a more diverse age demographic, consistent with the average age (30 years) in AR usability studies [22].

4.9 Display of Replication Study

The studies in our corpus used 29 different types of displays. The HTC Vive Pro was the most popular one, utilized in 21 studies (25% of all replication papers). The HTC Vive and Oculus Rift followed, with 12 studies (15%) and 6 studies (7%), respectively. In the AR domain, the HoloLens and its successor, the Microsoft HoloLens 2, were each used in 4 studies (5% of all replication papers). Other notable displays included the Oculus Quest 2, used in 3 studies (4%). Immersive projective systems (3), custom displays (2), and smartphones (1). 12 replication papers (14%) used the same display as the original studies.

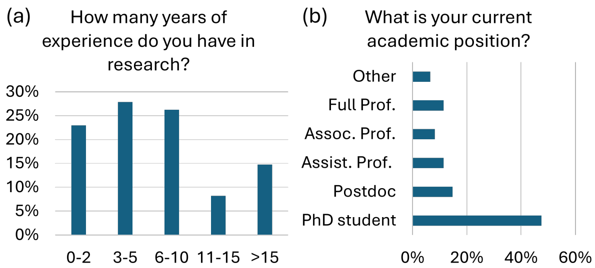

Figure 5: Demographics of the online survey participants, including years of research experience (a) and current academic position (b). "Other" includes 3 Master's students and 1 person from the industry.

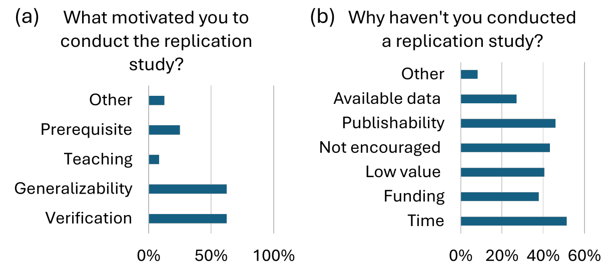

Figure 6: (a) Reasons that motivated researchers in our online survey to conduct a replication study. (b) Reasons for not conducting a replication study.

5 Online Survey

We complemented the literature survey with an online survey distributed among researchers in AR/VR. Through this, we gained additional insights into the kinds of experiences AR/VR researchers have had with replication studies and what encourages or discourages them from conducting replication studies.

5.1 Method

Participants in the online survey were recruited through mailing lists, social media, personal contacts, and attendees at the IEEE ISMAR conference. The ethics board of the University of Stuttgart approved the online survey. In the survey, participants were asked if they had previously replicated a study, what motivated them, or why they decided not to perform a replication study. The questions asking for reasons and motivation provided a selection of answers from which participants could choose, and the opportunity to provide additional reasons. In addition, they were asked if they had replicated one of their studies or studies of other researchers and whether they had access to artefacts from the prior study. Finally, they were asked how valuable they thought replication studies were and had the opportunity to provide additional comments.

5.2 Results

In this section, we present and discuss the insights gained from the online survey. The sample size was small, especially for certain experience levels, and participating researchers could be biased towards being more considerate of replication studies (self-selection bias). Therefore, these results should be interpreted with caution.

5.2.1 Demographics

61 participants completed the questionnaire. We obtained responses from participants with varying experience levels, ranging from less than two years to over 15 years. The distribution of the experience levels can be seen in Figure 5a. Similarly, the survey included participants with various academic positions, as shown in Figure 5b.

5.2.2 Why do people pursue replication research?

Overall, 39% (24) of the participants have conducted a replication study before. We observed the tendency that more experienced researchers are more likely to have conducted a replication study (0-2 years: 29%, 3-5 years: 29%, 6-10 years: 37%, 11-15 years: 80%, more than 15 years: 56%) which seems reasonable, yet should be taken with caution as our sample size is relatively small. 42% only replicated their studies, 29% only studies from other researchers, and 29% both kinds of studies. While only 33% of the participants with replication experience and up to 5 years of experience replicated their studies, 93% of participants with 6 or more years of experience have already replicated their studies.

The most common reasons for conducting a replication study, as mentioned by participants who had done replications, were to verify previous findings (63%) and to test generalizability in different settings (63%). Six participants (25%) also mentioned that it was a prerequisite for a larger project, and two participants (8%) did it for teaching purposes. One participant (1%) each also mentioned they wanted to compare it with their work and extend their experiment with prior research (cf. Figure 6a)).

Some participants also commented on why replication studies are important. They mentioned it is important because of fast-changing hardware (P11) and because some concepts have only been proposed but not properly evaluated (P13). P59 also mentions that replications would show if findings from the past are still valid and, therefore, should be cited, and P58 recommends doing replications before making major decisions based on prior study results.

5.2.3 Problems when pursuing replication research

When replicating studies, 64% of the participants mentioned having access to artefacts in addition to the article. However, several participants mentioned problems such as inconsistent information between published papers and corresponding dissertations (P57) and missing pieces (P47). In addition, P56 mentioned that the page limits make it impossible to add all the details of user studies. Another challenge identified by the participants was the non-random sampling in both the original and the replicated study (P49), which is intensified by publication bias, which prefers studies with high statistical differences that can often not be replicated (P58).

5.2.4 What prevents people from doing replication?

Participants who have not conducted a replication study so far explained this by practical issues such as a lack of time (51%), lack of funding (38%), and a lack of available data and other resources (27%). However, they also mentioned a low perceived value of replications (41%), that it is difficult to publish such studies (46%), and that the community does not encourage it (43%). An overview of these reasons is depicted in Figure 6b.

46% of participants who had not yet conducted a replication study mentioned that they at least considered doing so. Again, this seemed to be more common for more experienced researchers (0-2 years: 20%, 3-5 years: 42%, 6-10 years: 60%, 11-15 years: 100%, more than 15 years: 75%). Reasons for discarding the idea of conducting a replication study were similar to those mentioned before.

In the comment section, participants added that replication studies were often not valued, and, therefore, did not get published by large venues (P57) and would also not win awards or funding (P14). P2 mentioned "*a culture of importance among reviewers*" and that publishing a replication requires justifying the study's novelty besides the replication (P50). P13 experienced that even utilizing older methods might not be considered novel enough.

Several participants also expressed more fundamental arguments against replication studies, such as the concern of ending up in a rabbit hole of repeating the same studies (P34) and that most fundamental topics will be studied in different environments anyway, and, therefore, their validity will increase with time. P25 is also skeptical about the possibility of exact replications.

5.2.5 What is the perceived value of replications?

Participants rated their perceived value of replication studies at an average of 5.54 ($sd = 1.23$) on a scale from 1 (not valuable) to 7 (extremely valuable). Some participants mentioned conditions under which they find replication studies valuable, such as when comparing a new method to an existing one (P56) and if there is a valuable new aspect to the study (P30). P47 states that replications are important in fundamental AR/VR research, while context-specific problems are not worth replicating. Finally, P23 mentions that there should be a balance between replication and novelty.

5.2.6 Suggestions

In the open comment section, the participants also suggested guidelines such as providing more details in the paper or additional materials (P1), firmware and software, and recordings of the Unity experiments. P60 mentioned that it is important to consider details when replicating a study. To promote the publication of replication studies, P16 suggests encouraging young researchers to do replications, and P58 suggests that journals could fund replication studies as an incentive (they could also benefit from an increased impact factor). Finally, P47 comments that replications are crucial in the long run and "we quickly need to embrace and develop as a community, to allow for serious replications in the future."

6 General Discussion

Replication is critical to ensure findings remain robust and reliable in a rapidly evolving technological landscape. Our scoping literature review and accompanying survey reveal that replication studies in AR/VR are markedly scarce, with only 4% of all reviewed publications being replications, mirroring trends in related fields such as HCI and psychology. Moreover, it suggests that the XR community's attitude towards replication studies aligns with those in fields like HCI and psychology. Furthermore, this low prevalence underscores an ongoing challenge to verify foundational results while simultaneously pushing the boundaries of innovation.

Although replication studies in AR/VR are infrequent, the situation is complex and warrants careful analysis. Our analysis found no clear trend in who performed the replications. Original research teams often confirmed the stability of their findings, while independent groups provided evidence of generalizability across different settings and hardware. The majority of replications did not switch environments (e.g., from AR to VR) and used the same study design. In addition, our survey identified only 8 AR/VR topics for replication. However, by performing conceptual or partial replications, many expanded boundaries by changing the method or environment. This likely aims to test prior findings under novel conditions and increase perceived novelty over strict replications within the same environment.While such conceptual replications are important, they lead to a lack of strict replications. The lack of those (only 5 of 83) can be partially attributed to the rapid obsolescence of AR/VR hardware and the methodological complexities inherent to the systems. Our survey respondents also highlight publishability and perceived value as additional major barriers. Thus, some researchers try to go the middle way and perform partial or conceptual replications, strengthening prior findings while increasing the chances of their paper not getting rejected due to lack of novelty.

The participants from our online survey support that and suggest a community that values replication (5.54 on a scale from 1–7) but is constrained by practical issues such as resource limitations (e.g., funding, equipment, time) and cultural biases ("not encouraged"), favoring novel contributions over confirmatory work. However, the sample of our survey participants might be biased towards a more positive perception of replication studies. This tension between recognizing the importance of replication and the structural challenges that inhibit it forms a key area for understanding the current state of AR/VR research: Increasing the number of replications requires, therefore, cultural and institutional solutions.

6.1 A Single Replication?

The AR/VR community faces the same challenge as other domains: Very little research is replicated even once, let alone multiple times. At least in our corpus, only one original work was replicated more than once [100], and only three replications were replicated again. [73, 41, 53]. However, given findings from other domains, such as psychology [19], many apparent failed replications stem from methodological shortcomings rather than false original findings. For example, inflated effect sizes from selective reporting and data-driven analyses can mislead replication studies. In addition, replications may be underpowered, and the true effect may not be detected. Moreover, nonsignificant results in replications (and originals) may merely reflect the limitations of the applied statistical methods (e.g., insufficient sample size or violated assumptions). Alternative approaches, including Bayesian analysis, equivalence testing, meta-analysis, and multilevel modeling (account for the variance related to individuals in within-subjects designs), offer a fuller view of effect distributions. As a consequence, relying on a single replication is inadequate: Multiple replications of the same work would yield more reliable results. In addition, they could assess the influence of the samples' diverse international backgrounds (e.g., participants from South Korea compared to those from Finland) on the results. Thus, the AR/VR community should increase the number of replications in general and try to replicate findings multiple times. This requires that all stakeholders value, acknowledge, and promote replications, which need a shift in mindset.

6.2 Why are replications infrequent in AR/VR?

We identified the following core challenges:

Rapid Development of AR/VR Technology: The rapidly evolving nature of AR/VR technology presents significant hurdles for replication, particularly due to the constant updates of hardware and software platforms. Even when a replication study is attempted, it is often executed with updated devices and firmware (which might impact the 3D tracking or registration quality). As a result, researchers are frequently skeptical about the possibility of performing strict replications, which undermines efforts to confirm earlier findings.

Custom-made Display Availability: Examining the current custom-made displays in the AR community, we find that most are only available to particular labs. Those custom-made displays are unique, and their findings are novel. Replicating those findings with commonly available commercial AR displays is sometimes difficult or even impossible.

Methodological Challenges: Methodological constraints, such as limited access to detailed artefacts from original studies, inconsistent reporting practices, and the complexities of matching experimental setups, further deter replication efforts. These challenges are exacerbated in AR/VR, where user interactions and system complexities make precise replication inherently difficult.

Cultural Challenges: Culturally, the research community in AR/VR does not provide strong incentives for replication studies. The prevailing culture prioritizes novel contributions over confirmatory work. As our survey indicates, replication studies are perceived to have lower intrinsic value regarding career advancement, funding opportunities, and publication potential. This combines publication pressures that favor originality and novelty, leaving little room for studies that "simply" re-examine previous findings.

6.3 Recommendations

We propose recommendations to foster a more replication-friendly environment, thereby overcoming the challenges that hinder replication in AR/VR research. These recommendations address changes in research culture, improvements in replication reporting, and the establishment of better practices for replication studies. Moreover, we do not suggest that most studies need to undergo replication. However, the research community should maintain an optimal balance between exploratory and replication studies in a collective discovery effort.

6.3.1 Changes in Research Culture

One underlying challenge for replication work is the current focus on novelty by funders, publication venues, and during career advancement. In addition to focusing on novel findings, funding agencies, institutions, and publication venues must explicitly recognize and reward replication studies. Creating dedicated funding calls, awards, and publication tracks will help to align academic incentives with the need for verification. In addition, adopting open science practices, such as sharing data, software, and detailed experimental protocols, is key to making replication feasible. As seen in other disciplines, transparency and accessibility of research artifacts are critical for enabling other researchers to replicate and validate findings. Conferences and journals should encourage or mandate that authors submit high-quality supplemental material, enabling others to replicate the work. Training courses and workshops in empirical methods tailored for AR/VR research should be offered at all career stages. Educating early-career researchers on the importance and methodologies of replication can help instill a culture that values verification alongside innovation. In addition, they can be encouraged (or even required as part of a curriculum) to engage in replication work as a learning tool, while mid to senior career scholars can lead large-scale replication projects that build on and extend prior research (e.g., by combining a replication study with a follow-up question in a dedicated study leading to novel insights). Finally, academic evaluation criteria must embrace the importance of replication and value these works as an equally important part of an academic CV. By reducing the emphasis on novelty alone, the community can begin to appreciate the value of studies that confirm and generalize existing results.

6.3.2 Reporting of Replication

We recommend that the authors indicate when a study is a replication, specifying whether it is a strict, partial, or conceptual replication. This explicit labeling helps readers and reviewers understand the studys aims and context. To support the discoverability of replication studies, we recommend using the keyword section in papers, integrating specific replication categories in the submission systems, and potentially having dedicated issues (journals) and conference sessions. Authors should also provide comprehensive details of the original methods and highlight any deviations or adaptations made in the replication study. This includes precise information on hardware, software, experimental protocols, lab environment, and participant demographics. Reliance on supplementary materials or online repositories can alleviate space constraints.

6.3.3 Community Efforts

Replication studies, particularly in AR/VR, require significant resources in time, funding, and technical expertise. Many research groups may lack access to the necessary equipment or detailed artefacts from the original study, making it hard to replicate complex experimental conditions. Addressing these constraints should involve community-wide initiatives such as shared resource repositories or collaborative replication projects.

IEEE ISMAR and IEEE VR should develop and adopt protocols to conduct and report replication studies. Guidelines that outline best practices in experimental design, artefact sharing, and reporting can help ensure consistency and quality across studies.

With the rapid evolution of AR/VR hardware, establishing virtual archives or repositories for older devices can aid in conducting strict replications when needed. There is also a need for virtualization and simulation of (legacy) hardware. Such tools could mitigate the challenges posed by technological obsolescence. This is a promising avenue for future research.

7 Limitations and Future Work

Although we conducted the survey considering a large number of papers (2167 total papers) from 2010 to 2024 and gathered qualitative insights from researchers in AR/VR, our research has certain limitations that guide future replication research.

Our survey is limited to full papers of the top two major conferences (IEEE ISMAR and IEEE VR) in AR/VR research domains. Furthermore, our considered journal papers are only part of the conference-related IEEE TVCG Journal issues. We did not consider other issues of the IEEE TVCG from 2010 to 2024. Therefore, an extension of this survey could consider all issues from the IEEE TVCG journal and publications from other venues (e.g., ACM VRST, Journal of Virtual Reality, Frontiers in Virtual Reality) as well as posters and workshop papers. Moreover, our survey was not pre-registered.

Our online survey provided the results based on the responses of only 61 researchers, although we reached a wider range of people in the AR/VR research community. In addition, our online survey has a sampling bias (specifically, self-selection bias), as people who care about replications are more likely to respond to the survey than others. This could have influenced the results of the survey, especially regarding the reasons for not conducting replication studies. Also, the number of participants who conducted a replication before and the perceived value of replication studies might be lower in the overall research community. Future work should ensure a more complete sampling of participants.

8 Conclusions

This study presents a comprehensive examination of the state of replication research within the AR and VR domains, focusing on publications from the IEEE ISMAR and IEEE VR conferences over the past 15 years. Our findings reveal that replication studies constitute a mere 4% of the total publications, highlighting a significant gap in the empirical foundation of AR/VR research. Despite the recognized importance of replication in validating findings and ensuring scientific robustness, the AR/VR community has yet to fully embrace this practice. Through our scoping literature review and online survey of academics, we have identified several key barriers to conducting replication studies, including limited resources and cultural obstacles. These insights underscore the need for institutional support, policy changes, and dedicated platforms to foster a culture that values and encourages replication research. Our work contributes a publicly available collection of replication studies, qualitative insights into researchers' perspectives, and recommendations to increase the frequency of replication in AR/VR. By promoting higher research transparency and stronger validation of findings, we aim to enhance the reliability and efficiency of progress in AR/VR research. This study serves as a call to action for institutions, journals, conferences, and funding agencies to implement policies that support and incentivize replication, ultimately strengthening the scientific rigor of the field.

References

[1] F. Ahmed, J. D. Cohen, K. S. Binder, and C. L. Fennema. Influence of tactile feedback and presence on egocentric distance perception in virtual environments. In *IEEE VR*, 2010. doi: 10.1109/VR.2010.5444791 5

[2] V. Amrhein, D. Trafimow, and S. Greenland. Inferential Statistics as Descriptive Statistics: There Is No Replication Crisis if We Dont Expect Replication. *The American Statistician*, 2019. 2

[3] M. S. Arefin, N. Phillips, A. Plopski, J. L. Gabbard, and J. E. Swan. The effect of context switching, focal switching distance, binocular and monocular viewing, and transient focal blur on human performance in optical see-through augmented reality. *IEEE TVCG*, 2022. doi: 10.1109/TVCG.2022.3150503 4, 5

[4] A. U. Batmaz, M. D. B. Machuca, D. M. Pham, and W. Stuerzlinger. Do head-mounted display stereo deficiencies affect 3d pointing tasks in ar and vr? In *IEEE VR*, 2019. doi: 10.1109/VR.2019.8797975 5

[5] A. U. Batmaz, A. K. Mutasim, and W. Stuerzlinger. Precision vs. power grip: A comparison of pen grip styles for selection in virtual reality. In *IEEE VRW*, 2020. doi: 10.1109/VRW50115.2020.00012 5

[6] A. U. Batmaz and W. Stuerzlinger. Effective throughput analysis of different task execution strategies for mid-air fitts' tasks in virtual reality. *IEEE TVCG*, 2022. doi: 10.1109/TVCG.2022.3203105 5

[7] A. U. Batmaz, R. Turkmen, M. Sarac, M. D. B. Machuca, and W. Stuerzlinger. Effect of grip style on peripersonal target pointing in vr head mounted displays. In *IEEE ISMAR*, 2023. doi: 10.1109/ISMAR59233.2023.00057 5

[8] J. Baumeister, S. Y. Ssin, N. A. M. ElSayed, J. Dorrian, D. P. Webb, J. A. Walsh, T. M. Simon, A. Irlitti, R. T. Smith, M. Kohler, and B. H. Thomas. Cognitive cost of using augmented reality displays. *IEEE TVCG*, 2017. doi: 10.1109/TVCG.2017.2735098 5

[9] B. Benda, B. Rheault, Y. Lin, and E. D. Ragan. Examining effects of technique awareness on the detection of remapped hands in virtual reality. *IEEE TVCG*, 2024. doi: 10.1109/TVCG.2024.3372054 5

[10] L. Besançon and P. Dragicevic. The continued prevalence of dichotomous inferences at chi. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI EA '19. Association for Computing Machinery, New York, NY, USA, 2019. 2

[11] A. Bhargava, R. Venkatakrishnan, R. Venkatakrishnan, K. Lucaites, H. Solini, A. C. Robb, C. C. Pagano, and S. V. Babu. Can i squeeze through? effects of self-avatars and calibration in a person-plus-virtual-object system on perceived lateral passability in vr. *IEEE TVCG*, 2023. doi: 10.1109/TVCG.2023.3247067 5

[12] T. Blum, M. Wieczorek, A. Aichert, R. Tibrewal, and N. Navab. The effect of out-of-focus blur on visual discomfort when using stereo displays. In *2010 IEEE International Symposium on Mixed and Augmented Reality*, 2010. doi: 10.1109/ISMAR.2010.5643544 5

[13] B. Bodenheimer, S. Creem-Regehr, J. Stefanucci, E. Shemetova, and W. B. Thompson. Prism aftereffects for throwing with a self-avatar in an immersive virtual environment. In *IEEE VR*, 2017. doi: 10.1109/VR.2017.7892241 5

[14] L. E. Buck, S. Park, and B. Bodenheimer. Determining peripersonal space boundaries and their plasticity in relation to object and agent characteristics in an immersive virtual environment. In *IEEE VR*, 2020. doi: 10.1109/VR46266.2020.00053 5

[15] K. Caine. Local standards for sample size at chi. In *ACM CHI*, 2016. doi: 10.1145/2858036.2858498 6

[16] J. Casallas, J. Oliver, J. Kelly, F. Merienne, and S. Garbaya. Towards a model for predicting intention in 3d moving-target selection tasks. In *Engineering Psychology and Cognitive Ergonomics. Understanding Human Cognition*. Springer Berlin Heidelberg, 2013. doi: 10.1007/978-3-642-39360-0_2 5

[17] J. S. Casallas, J. H. Oliver, J. W. Kelly, F. Merienne, and S. Garbaya. Using relative head and hand-target features to predict intention in 3d moving-target selection. In *IEEE VR*, 2014. doi: 10.1109/VR.2014.6802050 5

[18] F. Chiossi, Y. Weiss, T. Steinbrecher, C. Mai, and T. Kosch. Mind the visual discomfort: Assessing event-related potentials as indicators for visual strain in head-mounted displays. In *IEEE ISMAR*, 2024. doi: 10.1109/ISMAR62088.2024.00029 5

[19] O. S. Collaboration. Estimating the reproducibility of psychological science. *Science*, 2015. doi: 10.1126/science.aac4716 2, 5, 6, 8

[20] C. A. T. Cortes, C.-T. Lin, T.-T. N. Do, and H.-T. Chen. An eeg-based experiment on vr sickness and postural instability while walking in virtual environments. In *IEEE VR*, 2023. 5

[21] G. Cumming. The new statistics: Why and how. *Psychological Science*, 2014. PMID: 24220629. 2

[22] A. Dey, M. Billinghurst, R. W. Lindeman, and J. E. Swan. A Systematic Review of 10 Years of Augmented Reality Usability Studies: 2005 to 2014. *Frontiers in Robotics and AI*, 2018. 2, 6

[23] T. D. Do, J. Benjamin, C. I. Protko, and R. P. McMahan. Cultural reflections in virtual reality: The effects of user ethnicity in avatar matching experiences on sense of embodiment. *IEEE TVCG*, 2024. doi: 10.1109/TVCG.2024.3456196 5

[24] P. Dragicevic. Fair Statistical Communication in HCI. In J. Robertson and M. Kaptein, eds., *Modern Statistical Methods for HCI*. Springer International Publishing, Cham, 2016. doi: 10.1007/978-3-319-26633-6_13 2

[25] S. Eom, D. Sykes, S. Rahimpour, and M. Gorlatova. Neurolens: Augmented reality-based contextual guidance through surgical tool tracking in neurosurgery. In *IEEE ISMAR*, 2022. doi: 10.1109/ISMAR55827.2022.00051 5

[26] S. A. Faleel, M. Gammon, K. Fan, D.-Y. Huang, W. Li, and P. Irani. Hpui: Hand proximate user interfaces for one-handed interactions on head mounted displays. *IEEE TVCG*, 2021. 5

[27] S. Fukushima, T. Hamada, and A. Hautasaari. Comparing world and screen coordinate systems in optical see-through head-mounted displays for text readability while walking. In *IEEE ISMAR*, 2020. doi: 10.1109/ISMAR50242.2020.00093 5

[28] J. L. Gabbard, D. G. Mehra, and J. E. Swan. Effects of ar display context switching and focal distance switching on human performance. *IEEE TVCG*, 2019. doi: 10.1109/TVCG.2018.2832633 4

[29] 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 *IEEE VR*, 2024. doi: 10.1109/VR58804.2024.00120 5

[30] H. C. Gagnon, T. Rohovit, H. Finney, Y. Zhao, J. M. Franchak, J. K. Stefanucci, B. Bodenheimer, and S. H. Creem-Regehr. The effect of feedback on estimates of reaching ability in virtual reality. In *IEEE VR*, 2021. doi: 10.1109/VR50410.2021.00107 5

[31] G. Gamelin, A. Chellali, S. Cheikh, A. Ricca, C. Dumas, and S. Otmane. Point-cloud avatars to improve spatial communication in immersive collaborative virtual environments. *Pers. Ubiq. Comput.*, 2020. doi: 10.1007/s00779-020-01431-1 5

[32] M. Gandy, R. Catrambone, B. MacIntyre, C. Alvarez, E. Eiriksdottir, M. Hilimire, B. Davidson, and A. C. McLaughlin. Experiences with an ar evaluation test bed: Presence, performance, and physiological measurement. In *IEEE ISMAR*, 2010. doi: 10.1109/ISMAR.2010.5643560 5

[33] H. Gao, K. Yue, S. Yang, Y. Liu, M. Guo, and Y. Liu. Exploring depth-based perception conflicts in virtual reality through error-related potentials. In *IEEE VR*, 2024. doi: 10.1109/VR58804.2024.00097 5

[34] D. Garrido, J. Jacob, and D. C. Silva. Performance impact of immersion and collaboration in visual data analysis. In *IEEE ISMAR*, 2023. doi: 10.1109/ISMAR59233.2023.00093 5

[35] O. S. Gómez, N. Juristo, and S. Vegas. Replications types in experimental disciplines. In *Proceedings of the 2010 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement*, ESEM '10. Association for Computing Machinery, New York, NY, USA, 2010. 3

[36] J. Gower, J. Baumeister, B. H. Thomas, and J. Zucco. Augmented reality annotations for assisting with decision-making and time-critical tasking. In *IEEE ISMAR*, 2024. doi: 10.1109/ISMAR62088.2024.00102 5

[37] C. Greiffenhagen and S. Reeves. Is replication important for hci? In *Proceedings of the CHI2013 Workshop on the Replication of HCI Research*, vol. 976, 2013. 2

[38] J. P. Grndling, D. Zeiler, and B. Weyers. Answering with bow and arrow: Questionnaires and vr blend without distorting the outcome. In *IEEE VR*, 2022. doi: 10.1109/VR51125.2022.00089 5

[39] R. Guski and N. Troje. Audiovisual phenomenal causality. *Perception & psychophysics*, 2003. doi: 10.3758/BF03194815 5, 6

[40] T. Hamada, A. Hautasaari, M. Kitazaki, and N. Koshizuka. Solitary jogging with a virtual runner using smartglasses. In *IEEE VR*, 2022. doi: 10.1109/VR51125.2022.00085 5

[41] D. Hayashi, K. Fujita, K. Takashima, R. W. Lindeman, and Y. Kitamura. Redirected jumping: Imperceptibly manipulating jump motions in virtual reality. In *IEEE VR*, 2019. doi: 10.1109/VR.2019.8797989 5, 8

[42] C. Hendrick. Replications, strict replications, and conceptual replications: Are they important? *Journal of Social Behavior & Personality*, 1990. 3

[43] T. Hoinville, A. Naceri, J. Ortiz, E. Bernier, and R. Chellali. Performances of experienced and novice sportball players in heading virtual spinning soccer balls. In *IEEE VR*, 2011. doi: 10.1109/VR.2011.5759441 5

[44] K. Hornbaek, S. S. Sander, J. A. Bargas-Avila, and J. Grue Simonsen. Is once enough?: on the extent and content of replications in human-computer interaction. In *ACM CHI*, 2014. doi: 10.1145/2556288.2557004 2, 3

[45] F. E. Ibrahim, H. H. Zayed, M. H. Koura, and N. ElSayed. Efficacy of virtual reality distraction for reducing chronic pain. In *IEEE ISMAR*, 2024. doi: 10.1109/ISMAR62088.2024.00119 5

[46] M. A. Jarrell and E. Peillard. Using identification with ar face filters to predict explicit & implicit gender bias. In *IEEE ISMAR*, 2023. doi: 10.1109/ISMAR59233.2023.00019 5

[47] S. Jeong, J. Kim, and J. Lee. The differential effects of multisensory attentional cues on task performance in vr depending on the level of cognitive load and cognitive capacity. *IEEE TVCG*, 2024. doi: 10.1109/TVCG.2024.3372126 5

[48] K. S. Jones, P. L. Derby, and E. A. Schmidlin. An Investigation of the Prevalence of Replication Research in Human Factors. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 2010. doi: 10.1177/0018720810384394 2

[49] S. Jung, Y. Wu, R. McKee, and R. W. Lindeman. All shook up: The impact of floor vibration in symmetric and asymmetric immersive multi-user vr gaming experiences. In *IEEE VR*, 2022. doi: 10.1109/VR51125.2022.00095 5

[50] J. W. Kelly, M. Hoover, T. A. Doty, A. Renner, M. Zimmerman, K. Knuth, L. A. Cherep, and S. B. Gilbert. Remote research on locomotion interfaces for virtual reality: Replication of a lab-based study on teleporting interfaces. *IEEE TVCG*, 2022. doi: 10.1109/TVCG.2022.3150475 4, 5, 6

[51] J. W. Kelly, A. G. Ostrander, A. F. Lim, L. A. Cherep, and S. B. Gilbert. Teleporting through virtual environments: Effects of path scale and environment scale on spatial updating. *IEEE TVCG*, 2020. doi: 10.1109/TVCG.2020.2973051 4, 6

[52] H. Khan and D. Nilsson. Smell of fire increases behavioural realism in virtual reality: A case study on a recreated mgm grand hotel fire. In *IEEE ISMAR*, 2023. doi: 10.1109/ISMAR59233.2023.00097 5

[53] E. M. Klose, N. A. Mack, J. Hegenberg, and L. Schmidt. Text presentation for augmented reality applications in dual-task situations. In *IEEE VR*, 2019. doi: 10.1109/VR.2019.8797992 5, 8

[54] D. Kodama, T. Mizuho, Y. Hatada, T. Narumi, and M. Hirose. Enhancing the sense of agency by transitional weight control in virtual co-embodiment. In *IEEE ISMAR*, 2022. doi: 10.1109/ISMAR55827.2022.00043 5

[55] H. Krist, E. L. Fieberg, and F. Wilkening. Intuitive physics in action and judgment: The development of knowledge about projectile motion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 1993. 6

[56] C. Krogmeier, E. Tison, J. Dillmann, A. Prouzeau, A. Prouteau, and M. Hachet. Leveraging augmented reality for understanding schizophrenia - design and evaluation of a dedicated educational tool. In *IEEE ISMAR*, 2024. doi: 10.1109/ISMAR62088.2024.00110 5

[57] B. Laha, D. A. Bowman, and J. D. Schiffbauer. Validation of the mr simulation approach for evaluating the effects of immersion on visual analysis of volume data. *IEEE TVCG*, 2013. doi: 10.1109/TVCG.2013.43 5

[58] B. Laha, K. Sensharma, J. D. Schiffbauer, and D. A. Bowman. Effects of immersion on visual analysis of volume data. *IEEE TVCG*, 2012. doi: 10.1109/TVCG.2012.42 5

[59] D. Lakens. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and anovas. *Frontiers in Psychology*, 2013. doi: 10.3389/fpsyg.2013.00863 2

[60] E. Langbehn, P. Lubos, G. Bruder, and F. Steinicke. Bending the Curve: Sensitivity to Bending of Curved Paths and Application in Room-Scale VR . *IEEE TVCG*, 2017. doi: 10.1109/TVCG.2017.2657220 5

[61] C. Lee, S. Bonebrake, D. A. Bowman, and T. Hllerer. The role of latency in the validity of ar simulation. In *IEEE VR*, 2010. doi: 10.1109/VR.2010.5444820 5

[62] L. S. Lenz, A. R. Fender, J. Chatain, and C. Holz. Comparing synchronous and asynchronous task delivery in mixed reality environments. *IEEE TVCG*, 2024. doi: 10.1109/TVCG.2024.3372034 5

[63] O. Li and H. Qiu. Virtual reality in supporting charitable giving: The role of vicarious experience, existential guilt, and need for stimulation. In *IEEE VR*, 2023. doi: 10.1109/VR55154.2023.00079 5

[64] Y. Li, I. A. Tahmid, F. Lu, and D. A. Bowman. Evaluation of pointing ray techniques for distant object referencing in model-free outdoor collaborative augmented reality. *IEEE TVCG*, 2022. doi: 10.1109/TVCG.2022.3203094 5

[65] J. Lin, J. Cronj, I. Kthner, P. Pauli, and M. E. Latoschik. Measuring interpersonal trust towards virtual humans with a virtual maze paradigm. *IEEE TVCG*, 2023. doi: 10.1109/TVCG.2023.3247095 5

[66] Y. Lin, T. Xie, Y. Chang, H. Luo, D. Wang, and S. Je. Vibroarm: Enhancing the sensation of forearm deformation in virtual reality using vibrotactile funneling illusion. In *IEEE ISMAR*, 2024. doi: 10.1109/ISMAR62088.2024.00066 5

[67] J. Liu, A. Jindal, C. Mantel, S. Forchhammer, and R. K. Mantiuk. How bright should a virtual object be to appear opaque in optical see-through ar? In *IEEE ISMAR*, 2022. doi: 10.1109/ISMAR55827.2022.00085 5

[68] J.-S. Liu, C. Elvezio, B. Tversky, and S. Feiner. Using multi-level precueing to improve performance in path-following tasks in virtual reality. *IEEE TVCG*, 2021. doi: 10.1109/TVCG.2021.3106476 5

[69] C. Lougiakis, A. Katifori, M. Roussou, and I.-P. Ioannidis. Effects of virtual hand representation on interaction and embodiment in hmd-based virtual environments using controllers. In *IEEE VR*, 2020. doi: 10.1109/VR46266.2020.00072 5

[70] A. MacQuarrie and A. Steed. Perception of volumetric characters' eye-gaze direction in head-mounted displays. In *IEEE VR*, 2019. doi: 10.1109/VR.2019.8797852 5

[71] M. R. Mahmud, M. Stewart, A. Cordova, and J. Quarles. Auditory feedback for standing balance improvement in virtual reality. In *IEEE VR*, 2022. doi: 10.1109/VR51125.2022.00100 5

[72] M. C. Makel, J. A. Plucker, and B. Hegarty. Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science*, 2012. PMID: 26168110. 2, 3

[73] A. Marquardt, C. Trepkowski, T. D. Eibich, J. Maiero, and E. Kruijff. Non-visual cues for view management in narrow field of view augmented reality displays. In *IEEE ISMAR*, 2019. doi: 10.1109/ISMAR.2019.000-3 5, 8

[74] A. Marquardt, C. Trepkowski, T. D. Eibich, J. Maiero, E. Kruijff, and J. Schning. Comparing non-visual and visual guidance methods for narrow field of view augmented reality displays. *IEEE TVCG*, 2020. doi: 10.1109/TVCG.2020.3023605 5

[75] A. Moraes, E. Hynes, R. Flynn, A. Hines, and N. Murray. Impact of auditory and audiovisual distractors on task performance in a vr-based auditory attention task. In *IEEE ISMAR*, 2024. doi: 10.1109/ISMAR62088.2024.00130 5

[76] A. N. Moraes, R. Flynn, A. Hines, and N. Murray. Evaluating the user in a sound localisation task in a virtual reality application. In *QoMEX*, 2020. doi: 10.1109/QoMEX48832.2020.9123136 5

[77] J. Neuliep and R. Crandall. Everyone was wrong: There are lots of replications out there. *Journal of Social Behavior & Personality*, 1993. 2, 3

[78] N. C. Nilsson, S. Serafin, and R. Nordahl. The effect of visual display properties and gain presentation mode on the perceived naturalness of virtual walking speeds. In *IEEE VR*, 2015. 5

[79] S. Nilsson, B. Johansson, and A. Jonsson. Using ar to support cross-organisational collaboration in dynamic tasks. In *IEEE ISMAR*, 2009. doi: 10.1109/ISMAR.2009.5336522 5

[80] B. A. Nosek and T. M. Errington. What is replication? *PLOS Biology*, 2020. doi: 10.1371/journal.pbio.3000691 1

[81] B. A. Nosek and D. Lakens. Registered reports: A method to increase the credibility of published results, 2014. 2

[82] K. Ogawa, K. Fujita, K. Takashima, and Y. Kitamura. Pseudojumpon: Jumping onto steps in virtual reality. In *IEEE VR*, 2022. doi: 10.1109/VR51125.2022.00084 5

[83] N. Ogawa, T. Narumi, and M. Hirose. Virtual hand realism affects object size perception in body-based scaling. In *IEEE VR*, 2019. doi: 10.1109/VR.2019.8798040 5

[84] R. Otono, A. Genay, M. Perusqua-Hernndez, N. Isoyama, H. Uchiyama, M. Hachet, A. Lcuyer, and K. Kiyokawa. I'm transforming! effects of visual transitions to change of avatar on the sense of embodiment in ar. In *IEEE VR*, 2023. 5

[85] P. Patil, R. Peng, and J. T. Leek. What Should Researchers Expect When They Replicate Studies? A Statistical View of Replicability in Psychological Science. *Perspectives on Psychological Science*, 2016. 2

[86] T. C. Peck, J. J. Good, and K. Seitz. Evidence of racial bias using immersive virtual reality: Analysis of head and hand motions during shooting decisions. *IEEE TVCG*, 2021. 5

[87] A. Peer and K. Ponto. Mitigating incorrect perception of distance in virtual reality through personalized rendering manipulation. In *IEEE VR*, 2019. doi: 10.1109/VR.2019.8797911 5

[88] T. Perrin, H. A. Kerherv, C. Faure, A. Sorel, B. Bideau, and R. Kulpa. Enactive approach to assess perceived speed error during walking and running in virtual reality. In *IEEE VR*, 2019. 5

[89] L. Phillips, B. Ries, M. Kaeding, and V. Interrante. Avatar self-embodiment enhances distance perception accuracy in non-photorealistic immersive virtual environments. In *IEEE VR*, 2010. doi: 10.1109/VR.2010.5444802 5

[90] E. Prytz, S. Nilsson, and A. Jnsson. The importance of eye-contact for collaboration in ar systems. In *IEEE ISMAR*, 2010. doi: 10.1109/ISMAR.2010.5643559 5

[91] A. Raikwar, D. Mifsud, C. D. Wickens, A. U. Batmaz, A. C. Warden, B. Kelley, B. A. Clegg, and F. R. Ortega. Beyond the wizard of oz: Negative effects of imperfect machine learning to examine the impact of reliability of augmented reality cues on visual search performance. *IEEE TVCG*, 2024. doi: 10.1109/TVCG.2024.3372062 5

[92] G. Richard, T. Pietrzak, F. Argelaguet, A. Lcuyer, and G. Casiez. Within or between? comparing experimental designs for virtual embodiment studies. In *IEEE VR*, 2022. 5

[93] M. Rietzler, J. Gugenheimer, T. Hirzle, M. Deubzer, E. Langbehn, and E. Rukzio. Rethinking redirected walking: On the use of curvature gains beyond perceptual limitations and revisiting bending gains. In *IEEE ISMAR*, 2018. doi: 10.1109/ISMAR.2018.00041 5

[94] C. S. Rosales, G. Pointon, H. Adams, J. Stefanucci, S. Creem-Regehr, W. B. Thompson, and B. Bodenheimer. Distance judgments to on- and off-ground objects in augmented reality. In *IEEE VR*, 2019. doi: 10.1109/VR.2019.8798095 5

[95] R. Rzayev, P. W. Woźniak, T. Dingler, and N. Henze. Reading on smart glasses: The effect of text position, presentation type and walking. In *ACM CHI*, 2018. doi: 10.1145/3173574.3173619 5

[96] J. Shen, J. J. Dudley, and P. O. Kristensson. Encode-store-retrieve: Augmenting human memory through language-encoded egocentric perception. In *IEEE ISMAR*, 2024. 5

[97] S. Sigurdarson, A. P. Milne, D. Feuereissen, and B. E. Riecke. Can physical motions prevent disorientation in naturalistic vr? In *IEEE VRW*, 2012. doi: 10.1109/VR.2012.6180874 5

[98] U. Simonsohn, L. Nelson, and J. Simmons. P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General*, 2014. 2

[99] A. D. Souchet, M. L. Diallo, and D. Lourdeaux. Cognitive load classification with a stroop task in virtual reality based on physiological data. In *IEEE ISMAR*, 2022. 5

[100] F. Steinicke, G. Bruder, J. Jerald, H. Frenz, and M. Lappe. Estimation of detection thresholds for redirected walking techniques. *IEEE TVCG*, 2009. 8

[101] J. E. Swan II and J. L. Gabbard. Survey of user-based experimentation in augmented reality. In *Proceedings of 1st International Conference on VR, HCI International*, 2005. 2

[102] R. Todd, Q. Zhu, and A. Bani. Temporal availability of ebbinghaus illusions on perceiving and interacting with 3d objects in a contextual virtual environment. In *IEEE VR*, 2021. 5

[103] Á. Torres, J. P. Molina, A. S. Garca, and P. Gonzlez. Prototyping of augmented reality interfaces for air traffic alert and their evaluation using a virtual reality aircraft-proximity simulator. In *IEEE VR*, 2024. doi: 10.1109/VR58804.2024.00101 5

[104] A. C. Tricco, E. Lillie, W. Zarin, K. K. O'Brien, H. Colquhoun, D. Levac, D. Moher, M. D. Peters, T. Horsley, L. Weeks, et al. Prisma extension for scoping reviews (prisma-scr): Checklist and explanation. *Annals of Internal Medicine*, 2018. PMID: 30178033. doi: 10.7326/M18-0850 2

[105] R. Venkatakrishnan, R. Venkatakrishnan, B. Raveendranath, R. Canales, D. M. Sarno, A. C. Robb, W. Lin, and S. V. Babu. The effects of secondary task demands on cybersickness in active exploration virtual reality experiences. *IEEE TVCG*, 2024. doi: 10.1109/TVCG.2024.3372080 5, 6

[106] R. Venkatakrishnan, R. Venkatakrishnan, B. Raveendranath, D. M. Sarno, A. C. Robb, W.-C. Lin, and S. V. Babu. The effects of auditory, visual, and cognitive distractions on cybersickness in virtual reality. *IEEE TVCG*, 2023. 6

[107] B. Volmer, J. Baumeister, R. Matthews, L. Grosser, S. Von Itzstein, S. Banks, and B. H. Thomas. A comparison of spatial augmented reality predictive cues and their effects on sleep deprived users. In *IEEE VR*, 2022. doi: 10.1109/VR51125.2022.00079 5

[108] E.-J. Wagenmakers, R. Wetzels, D. Borsboom, H. L. J. van der Maas, and R. A. Kievit. An agenda for purely confirmatory research. *Perspectives on Psychological Science*, 2012. PMID: 26168122. 2

[109] D. Wang, J. Kubrlcht, Y. Zhu, W. Lianq, S.-C. Zhu, C. Jiang, and H. Lu. Spatially perturbed collision sounds attenuate perceived causality in 3d launching events. In *IEEE VR*, 2018. 5, 6

[110] F. Weidner, J. E. Maier, and W. Broll. Eating, smelling, and seeing: Investigating multisensory integration and (in)congruent stimuli while eating in vr. *IEEE TVCG*, 2023. 5

[111] C. Wienrich, R. Gross, F. Kretschmer, and G. Mller-Plath. Developing and proving a framework for reaction time experiments in vr to objectively measure social interaction with virtual agents. In *IEEE VR*, 2018. doi: 10.1109/VR.2018.8446352 5

[112] M. L. Wilson, E. H. Chi, D. Coyle, and P. Resnick. Chi workshop on the replication of hci research (replichi). Proceedings of the ACM CHI Workshop on the Replication of HCI Research, 2013. 2

[113] M. L. Wilson, W. Mackay, E. Chi, M. Bernstein, D. Russell, and H. Thimbleby. RepliCHI - CHI should be replicating and validating results more: discuss. In *CHI Extended Abstracts*, 2011. 2

[114] E. Wolf, N. Dllinger, D. Mal, C. Wienrich, M. Botsch, and M. E. Latoschik. Body weight perception of females using photorealistic avatars in virtual and augmented reality. In *IEEE ISMAR*, 2020. doi: 10.1109/ISMAR50242.2020.00071 5

[115] T. Ye, S. Qi, J. Kubricht, Y. Zhu, H. Lu, and S.-C. Zhu. The martian: Examining human physical judgments across virtual gravity fields. *IEEE TVCG*, 2017. doi: 10.1109/TVCG.2017.2657235 5, 6

[116] T. Yin, L. Hoyet, M. Christie, M.-P. Cani, and J. Pettr. The one-man-crowd: Single user generation of crowd motions using virtual reality. *IEEE TVCG*, 2022. doi: 10.1109/TVCG.2022.3150507 5

[117] B. Yoon, H.-i. Kim, G. A. Lee, M. Billinghurst, and W. Woo. The effect of avatar appearance on social presence in an augmented reality remote collaboration. In *IEEE VR*, 2019. 5

[118] C. You, B. Benda, E. S. Rosenberg, E. Ragan, B. Lok, and J. Thomas. Strafing gain: Redirecting users one diagonal step at a time. In *IEEE ISMAR*, 2022. doi: 10.1109/ISMAR55827.2022.00077 5

[119] K. Yu, G. Gorbachev, U. Eck, F. Pankratz, N. Navab, and D. Roth. Avatars for teleconsultation: Effects of avatar embodiment techniques on user perception in 3d asymmetric telepresence. *IEEE TVCG*, 2021. doi: 10.1109/TVCG.2021.3106480 5

[120] A. Zenner and A. Krger. Estimating detection thresholds for desktop-scale hand redirection in virtual reality. In *IEEE VR*, 2019. 5