Stay Vigilant: The Threat of a Replication Crisis in VR Locomotion Research

Daniel Zielasko*
University of Trier
Trier, Germany
daniel.zielasko@rwth-aachen.de

Tim Weissker*
RWTH Aachen University
Aachen, Germany
me@tim-weissker.de

ABSTRACT

The ability to reproduce previously published research findings is an important cornerstone of the scientific knowledge acquisition process. However, the exact details required to reproduce empirical experiments vary depending on the discipline. In this paper, we summarize key replication challenges as well as their specific consequences for VR locomotion research. We then present the results of a literature review on artificial locomotion techniques, in which we analyzed 61 papers published in the last five years with respect to their report of essential details required for reproduction. Our results indicate several issues in terms of the description of the experimental setup, the scientific rigor of the research process, and the generalizability of results, which altogether points towards a potential replication crisis in VR locomotion research. As a countermeasure, we provide guidelines to assist researchers with reporting future artificial locomotion experiments in a reproducible form.

CCS CONCEPTS

• **Human-centered computing** → **Virtual reality**; **Empirical studies in HCI**; • **General and reference** → *Surveys and overviews.*

KEYWORDS

Virtual Reality, Reproducibility, Replication Crisis, Locomotion, Steering, Teleportation

ACM Reference Format:
Daniel Zielasko and Tim Weissker. 2023. Stay Vigilant: The Threat of a Replication Crisis in VR Locomotion Research. In *29th ACM Symposium on Virtual Reality Software and Technology (VRST 2023), October 9–11, 2023, Christchurch, New Zealand*. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3611659.3615697

1 INTRODUCTION

Replication is a core principle of empirical sciences because we can never fully capture nor control our environment with all its variables. It is through the repeated observation of identical effects on the same or similar stimuli that we can attain reliable knowledge. However, there is an increasing concern that a large proportion of published research results are not replicable. A collaborative effort in psychological science, for instance, attempted the replication of 100 psychological studies and found that, while 97 % of original studies had significant results, the same was true for only 37% of the replications. Moreover, the mean effect size of the replication effects was half the magnitude of the mean effect size observed in the original studies [27]. While such a large project has not yet been conducted in computer science, researchers have pointed towards the risk of similarly poor success rates [11]. However, a review of papers in 2014 covering the broad field of human-computer interaction (HCI) revealed that replication studies are uncommon as they were contained in only 3% of the surveyed papers [18]. As a consequence, the extent of a potential replication crisis in HCI research is currently unclear.

To approach this gap, this paper focuses on the reproducibility of locomotion experiments in virtual reality (VR) as a specific subarea of HCI research. These experiments are especially interesting to analyze as the reporting of 3D user interfaces and their evaluations often encounter significant challenges due to the technical intricacies involved. This paper contributes a formal literature review in an attempt to quantify the reproducibility of different VR locomotion experiments with respect to three central facets. First, we investigated whether experimental setups were described in sufficient detail such that they can be reconstructed by an experienced practitioner. This is important as the replication of experimental results is impossible if the experiment that led to these results cannot be reconstructed. Second, we studied whether standardized empirical evaluation procedures like null hypothesis significance testing (NHST) were conducted and reported with the required scientific rigor that these procedures demand. If done correctly, this increases the likelihood of producing correct results and conclusions that therefore have a higher chance of reproduction. Finally, we analyzed the reported samples to see if the presented results are likely to be generalizable. This is relevant as scientific findings are more likely to be reproducible when they apply to the general population rather than a specific subset of it. Based on these three facets of reproducibility, our overarching research question of this paper was as follows: **Does state-of-the-art research in VR locomotion fulfill basic requirements for reproducibility?**

As an initial approach towards answering this question, our literature review presented in this paper focuses on artificial locomotion techniques [7], including steering and teleportation. We will use the two terms as follows. Teleportation is characterized as instantaneous (egocentric) viewpoint modification through target-, less common direction-, specification [30]. Steering includes virtual and continuous locomotion techniques with continuous direction specification and usually also continuous speed specification [5, 8], i.e., a continuous specification of viewpoint velocity. Both paradigms are among the most common when it comes to travel interfaces in virtual reality applications [23]. The limitation to these two paradigms allows us to set a stronger focus on their specific characteristics

*Both authors contributed equally to this research.

VRST 2023, October 9–11, 2023, Christchurch, New Zealand

and therefore provide specific guidelines on writing reproducible descriptions of these.

In summary, we make the following scientific contributions:

- we summarize important challenges for reproducibility and their specific practical implications for VR locomotion research
- we conduct a formal literature review that clearly points towards the threat of a potential replication crisis in VR locomotion research
- we present an overview of existing experimental design and evaluation practices in the domain of locomotion research
- we formulate specific guidelines that help to report experiments in locomotion research, especially for steering and teleportation interfaces

2 REPLICATION CHALLENGES IN VR LOCOMOTION RESEARCH

The successful reproduction of scientific results in the realm of VR locomotion techniques depends on a large variety of factors on both the side of the original authors as well as the scientists attempting the reproduction. In this paper, our focus is on the design and presentation of the original experiments in their respective publications, where we put a particular emphasis on three underlying key aspects that directly influence the chances of a successful replication by other scientists. These are the description of the experimental setup (Section 2.1), the scientific rigor of the research process (Section 2.2), and the generalizability of the results (Section 2.3). In the following, we will underline the importance of each aspect and explain the most relevant resulting challenges in the context of VR locomotion research.

2.1 Description of Experimental Setup

Experimental results can only be reproduced when sufficient and accurate details are provided on how the experiment was designed, conducted, and evaluated. This typically includes an unambiguous description of the employed hard- and software, the experimental structure and procedure, the sample and population from which it was generated, and the data analysis on which the overarching conclusions are based. While these considerations are obvious in theory, there are practical challenges specific to VR locomotion research that stand in the way of their realization.

First, different parametrizations of the same locomotion techniques and even different locomotion techniques are often referenced by the same name. *Leaning*, as a representative of *Steering* interfaces, refers to both stepping and balancing interfaces [25] as well as interfaces that involve inclining different parts of the body, from the user's head [43] to the entire body [24]. The corresponding inclination angle can be either measured directly [44] or more commonly derived from position offsets [25, 43] for technical reasons. *Teleportation*, on the other hand, can refer to long-distance travel via an overview representation of the world [15, 40] as well as short-distance travel within the currently visible part of the scene [30] while both may use entirely different types of transition to relocate the user [1, 30, 40]. In addition, plugins for major game engines that are the basis of many research prototypes often do not use the same default settings, further blurring the common ground for discussions. As a result, reporting the alleged name of a technique is often not enough to convey how it was implemented.

Second, the established publication format of a multi-page essay is not well-suited for the presentation of interactive and time-variant experiences such as VR interfaces. Therefore, readers often need to infer relevant details when translating the given texts and figures into a mental image of the experience.

Finally, different locomotion techniques were previously shown to have an influence on several human-related factors that are also strongly affected by other variables. A large body of research, for example, analyzes the effects of virtual locomotion on spatial awareness (e.g., [32, 40, 42]), a complex cognitive construct that was shown to vary based on age [38], gender [10], and spatial activities as a child [14]. Others focus on the consequences of locomotion on cybersickness with similarly large inter- and intrapersonal differences [12, 33, 41]. Furthermore, the learning curve and overall performance of individual participants in studies on 3D user interfaces are both directly related to their prior experience with similar interfaces. Therefore, providing an accurate description of participant samples including all relevant factors that might influence the results produced by these participants is challenging, which is aggravated by the fact that research regarding several of these factors is still ongoing.

2.2 Scientific Rigor of the Research Process

Producing meaningful research findings relies on the rigorous application of scientific methods to keep the rate of false positive or false negative results low. While the exact procedure varies across methods, all of them typically start with initial observations, ideas, or beliefs that shape the formulation of one or multiple research question(s) [11, 17]. Based on our experiences in VR locomotion research, the formulation of these questions can already pose a challenge since one has to clearly distinguish between aspects relating to the technical realization of locomotion techniques (typically answered by accurate hard- and software descriptions as introduced before) as opposed to aspects relating to the effects that locomotion techniques have on their users (typically answered by the results of an empirical experiment). In this paper, we further focus on null hypothesis significance testing (NHST) as the most prevalent framework for drawing conclusions from empirical results in VR locomotion research, which is however increasingly criticized for its risk of misinterpreting results due to errors in its rigorous application (e.g., [3, 11, 21, 39]). We would therefore like to summarize two important pillars of a rigorous application of NHST that are essential for meaningful conclusions in VR locomotion research and beyond. First, NHST requires the formulation of concrete hypotheses to be evaluated, the desired significance level (α), and the required number of participants before starting with the data collection. Changing any of these parameters during or after the experiment can manipulate experimental conclusions due to a violation of the underlying statistical assumptions of NHST [11, 22]. Second, the results of the corresponding statistical tests need to be reported completely, including negative results, and given in precise numbers to reflect the statistical uncertainty that comes with their application [34, 39]. This practice also supports meta-analyses and power analyses (see below) in subsequent work.

Several tools and frameworks have been proposed to assist with a rigorous application of NHST. For example, online platforms like OSF[1] allow for the pre-registration of hypotheses to improve research transparency. Furthermore, calculators like G*Power[2] conduct statistical power analyses to determine an appropriate number of participants based on the expected effect size. In VR locomotion research, however, it is often challenging to tell the expected effect size in advance, especially for novel technique proposals.

2.3 Generalizability of Results

Empirical studies aim to draw general conclusions about the entirety of a certain population by looking at the data gathered from a limited sample. To allow for a successful replication of these inferences, it is crucial that the sample is a random, sufficiently large, and representative selection of members from the population of interest. While VR locomotion studies inherit these general challenges, it is especially important to consider and discuss to which population the obtained results are likely to apply. Beyond the commonly cited concern that many studies in HCI and VR only sample from a young, academic, and predominantly male population (e.g. [2, 13, 26, 29]), calls for participation in VR locomotion studies are, based on our experiences, most frequently answered by tech-savvy people who sometimes even own VR hardware themselves. While gathering feedback from such a proficient user base can be particularly insightful for developing expressive expert interfaces, the same results might not apply to a more general population.

Moreover, it is important to note that locomotion in VR is often only a sub-task of a more high-level objective [6]. Beyond the exact replication of a specific experiment, it is therefore also desirable to confirm the effects of a locomotion technique in different task setups. However, obtaining novel results in such a replication attempt poses the additional challenge of distinguishing between the influences of the locomotion technique and the task setup. To do so, further additional user studies may be required.

2.4 Discussion

While certainly not an exhaustive list, the aforementioned examples demonstrate that the replication of VR locomotion experiments faces general challenges shared with many other disciplines while adding more specific concerns that all have to be addressed properly. It is therefore crucial that VR locomotion researchers are aware of the replication issues in their field and try to prevent them from the very beginning of their research process. Our literature review reported in this paper provides an overview of the current state of several replication issues in research papers published in the previous five years. Our detailed methodology as well as results will be presented in the following sections.

3 LITERATURE REVIEW OF VR LOCOMOTION STUDIES

To approach our research question of whether current practices regarding the design and presentation of VR locomotion experiments allow for reproduction, we conducted a formal literature review of 61 locomotion papers published in the previous five years.

3.1 Inclusion Criteria

Our survey was focused on full papers that appeared between January 2018 and June 2023 at one of three major conferences focused on publishing research on virtual reality. The included venues were the IEEE Conference on Virtual Reality and 3D User Interfaces, the IEEE International Symposium on Mixed and Augmented Reality, and the ACM Symposium on Virtual Reality Software and Technology as well as invited journal articles from IEEE Transactions on Visualization and Computer Graphics (TVCG) presented at these venues.

To identify publications in the realm of VR locomotion research, we manually went through the titles and abstracts of all publications and scanned them for the keywords *navigation*, *travel*, *locomotion*, *viewpoint manipulation*, *steering*, *teleportation* as well as keywords with synonymous meaning. The remaining papers were read in detail to identify the ones that contained at least one empirical user study. To further reduce the resulting collection, we then decided to focus on studies that involved at least one *artificial* locomotion method (i.e., steering or teleportation) in the typology of Boletsis and Chasanidou [4]. We, therefore, implicitly excluded papers studying only real or redirected walking techniques (*physical* locomotion) as well as papers that developed machine-learning models without accompanying empirical user studies.

After this process, we arrived at a final collection of 61 publications (including 15 TVCG journal papers) that formed the basis for our evaluations carried out in this paper. The full list of papers is provided as supplemental material together with this article.

3.2 Annotation Procedure

After the initial selection process, the 61 papers were once again read in detail with a particular focus on reproducibility. For this purpose, we extracted information on a selection of attributes that we considered relevant for understanding and reproducing the presented experiments based on our discussions provided in Section 2. In addition to the binary and ternary attributes summarized in Table 1, these also included numeric and categorical information on the presented techniques, the type of study design, the formulated and accepted hypotheses, the demographics of participant samples, and the statistical analysis methods used. Since the goal of this paper is to gain an overarching impression of reproducibility issues in VR locomotion research rather than pointing out the flaws of individual papers, we then aggregated our results into descriptive statistics representing the entire collection. In the following section, we will go through all of the collected attributes individually and report on our corresponding findings.

4 RESULTS

After starting with the report of a few general observations and findings, our results are structured by the three aspects of reproducibility introduced in Section 2.

[1]https://osf.io/
[2]https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower

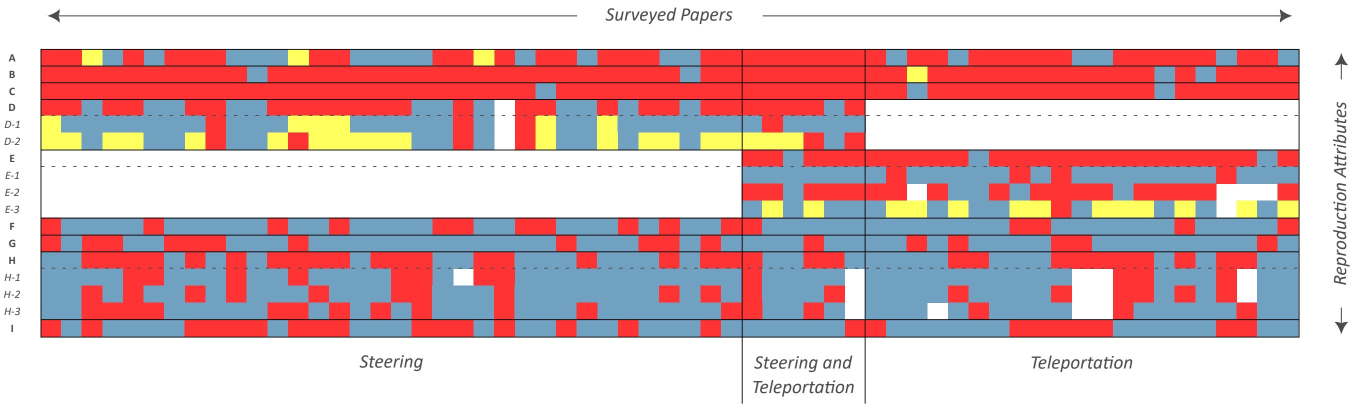

Figure 1: Overview of our selected binary and ternary attributes (rows) as well as their values for each of the 61 surveyed papers mapped to colors (columns). While every red box indicates a missing piece of information, yellow boxes indicate an ambiguous or incomplete piece of information. Blue boxes show that the corresponding piece of information was provided. White boxes indicate that an attribute did not apply to that paper. A description of the visualized attributes is given in Table 1.

Table 1: Descriptions of the binary and ternary attributes used for analyzing relevant aspects of reproducibility. The letters refer to the attribute rows in the visualization of our data set in Figure 1.

Attribute	Description
A	Explicit mention of a research question
B	Availability of source code
C	Availability of an OSF project
D	Completeness of steering report (*aggregate of sub-attributes*)
D-1	Report of direction specification
D-2	Report of transfer function for speed specification
E	Completeness of teleportation report (*aggregate of sub-attributes*)
E-1	Report of target specification function
E-2	Report of target specification reach
E-3	Report of transition type
F	Description of population or sampling strategy
G	Report of participants' prior VR experience
H	Complete NHST reports *(aggregate of sub-attributes)*
H-1	Report of exact p-values
H-2	Report of test statistics
H-3	Report of non-significant results
I	Report of effect sizes

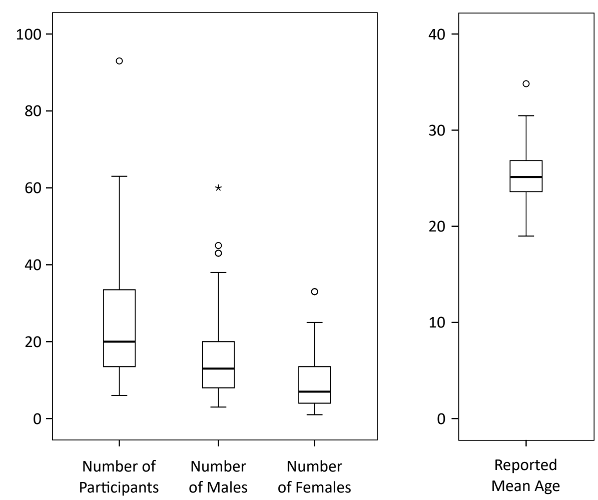

Figure 2: Boxplots illustrating the distribution of participant counts divided by gender in all surveyed user studies (N = 89) as well as the distribution of reported mean ages in studies where this information was provided (N = 66).

4.1 General Findings

Of the 61 papers in our collection, 40 thematized steering while 27 thematized teleportation. As a result, six papers addressed both steering and teleportation techniques. All papers described a total of 89 studies, which results in an average of 1.46 (SD = 0.70, Mdn = 1) studies per paper. 71 of these studies employ a within-subjects design, 12 a between-subjects design, 3 a mixed methods design, and 3 run only one condition. The mean number of participants is 25.1 (SD = 15.2, Mdn = 20), which is much higher than the most common sample size of 12 found at the ACM CHI conference [9] and the 10 ± 2 rule for usability evaluations derived by Hwang et al. in 2010 [20]. However, it is discussed that this rule also is too general and requires more context than simply "usability research" [35]. A visual overview of the distributions of participant counts per gender as well as the reported mean ages is given in Figure 2.

4.2 Description of Experimental Setup

For the reproducibility of the experimental setup, we looked at two variables as examples. First, "*is the work presented reproducible by an experienced practitioner/researcher in the area?*" (reviewing form, IEEE VR 2023). And second, *is the study population and sample* *described in such a way that they can be replicated?* In the following, we test this with a general set of information that is necessary for the replication of the experiment by an uninvolved third person, but not necessarily sufficient. The experimental designs might require the report of additional variables that are not captured here.

4.2.1 Steering. For the steering interfaces, we consider 13 out of 39 reproducible (1 does not suit the used scheme), which is 33% and summarized in Row D of Figure 1. More concretely, we look at the specification of the direction component and the specification of the motor/speed component. We consider an interface not reproducible if at least one of these components was not specified or specified with ambiguities. 29 of the papers report the direction specification sufficiently, 6 with ambiguities, and 4 not at all. 13 of the papers report the speed component sufficiently, 20 with ambiguities, and 6 not at all. Ambiguities most often arise from missing parameters, specification of a maximum speed without the interpolating function (including non-specified deadzones), or the (implicit) assumption of some non-existing standard (made up for illustration: *"we used a gamepad interface"*).

4.2.2 Teleportation. For the teleportation interfaces, we consider 3 out of 27 reproducible, which is 11% and summarized in Row E of Figure 1. More concretely, we look at the type of target specification (e.g., parabola or straight ray), the maximum target specification distance (i.e., the reach of the teleport), and the specification of the transition (e.g., instantaneous, fade-to-black over x seconds, etc.). We consider an interface not reproducible if at least one of these components was not specified or specified with ambiguities. 23 of the papers report the target specification type sufficiently and 4 do not specify it at all. 5 of the papers report the maximum reach of the selection tool, 18 do not, and for 3 this variable does not apply. 12 of the papers report the transition sufficiently, 13 with ambiguities, and 1 not at all. Furthermore, we observe but do not quantify that visual mediators, such as preview avatars or spatial orientation helpers like arrows, are often not reported in detail. Finally, it is often not specified whether the interfaces include an option for virtual rotation.

Across both, steering and teleportation interfaces, 5 of the 61 analyzed papers publish their software along with the paper, 4 of them in the form of the source code and 1 in the form of standalone executables.

4.2.3 Population & Sample. For 65 out of 89 studies, the authors at least sketch the population that their participants' sample was drawn from (e.g., social media or university background, see Section 4.4). An explicit sampling strategy, such as *convenience* or *stratified* sampling, is not reported in any of the papers. For the further specification of the sample, we concentrate on the reporting of the distribution of gender, age, and prior VR experience.

83 studies report a gender distribution and 6 do not (see Section 4.4 for more details).

For 86 of the studies, some description of the participant's ages is given, no information is provided for 3 studies, and 32 studies lack a measure of variance. This means that for the majority of the studies, the age of the participants is described in some form, which however varies across publications. In particular, 66 cases report the mean age (grand mean of mean ages: 25.4, SD = 2.7, range: 19-34.83), 56 cases report the range (from overall 12 to 66), and 8 cases report the median (grand mean of median ages: 24.2, SD = 2.6, range: 21-27.5). The standard deviation is used as the measure of dispersion in 65 cases (grand mean of standard deviations: 4.6, SD = 2.2, range: 1.29-13.2).

Most of the study reports, 69 out of 89, also further specify the sample's prior experience with VR. However, the variance of the chosen measure is even greater here:

- 26 provide a boolean value (e.g., had prior experience)
- 13 provide general user categories (e.g., beginner, expert)
- 11 provide Likert scale ratings (e.g. from 1 to 7)
- 7 provide prior experiences in a related discipline (e.g., 3D graphics or video)
- 5 provide absolute prior usage times (e.g. 5h used so far)
- 5 provide usage regularity (e.g. every week, every month)
- 2 provide a time span of familiarity (e.g. used HMDs for more than x years)

An additional observation is that among the 15 studies that use a between-subjects design, only 3 report the distribution of participants within each individual group.

4.3 Scientific Rigor of the Research Process

While the prior section focused on the reproducibility of the underlying experiment, we now take a look at other reasons why experiments potentially cannot be reproduced or compared with other results. As discussed in Section 2.2, the basis of many discussions about the use of NHST is that the process is often not followed rigorously, i.e., that conclusions are drawn without having first posed questions and hypotheses or that only specific (often significant) results are reported.

Contrary to the ongoing discussion of its pitfalls, the overwhelming majority of all work, namely 57 out of 61, still relies mainly on NHST. Only 3 of the papers rely on qualitative analysis, while 1 concentrates on the reporting of confidence intervals and effect sizes. Therefore, we focus our further analysis mainly on the rigorous process in NHST reporting. The following descriptive statistics, however, do not include studies that were explicitly flagged as prestudies by the authors.

4.3.1 Research Questions & Hypotheses. 19 of the 61 analyzed papers explicitly formulate one or more research questions. 42 papers do not specify an explicit research question. The requirement of mentioning an explicit research question is very strict since the scientific question can often be derived from the context or the experiment, even though this requires more effort for the reader and potentially raises ambiguities.

39 of 61 (35 of 57 that use NHST) papers specify at least one hypothesis (mean number of hypotheses: 3.9, SD = 2.1, Mdn = 4). However, 1 of these papers does not test their hypotheses. 11 papers accept hypotheses partially. 3 of 61 papers chose to preregister their hypotheses and parts of the study design using OSF.

4.3.2 Statistical Reporting. Regarding the reporting of the inferential statistics, 30 of 61 papers report their statistical results completely, which is 49%. More precisely, this breaks down into the following three subcategories: 41 out of 60 do report non-significant results and do not leave out their exact values. 44 out of 56 report exact p-values. 39 out of 57 do report other statistical results in detail, such as F-statistics, etc. The total numbers above are different from 61 because no statement can be made for the remaining papers, for example, in cases where there were no non-significant results. Finally, 36 out of the 61 papers report at least one effect size.

4.4 Generalizability of Results

Finally, we argued that reproducibility also relates to the applicability of findings beyond the samples of the user studies. The information provided in the surveyed works allows us to gain an initial impression of whether such a generalization might be possible. Across all studies that report gender distributions, 63% and 37% of participants are male and female, respectively. More specifically, the mean number of female participants per study is 9.3 (SD = 7.0, Mdn = 7) while the mean number of male participants is 15.8 (SD = 10.7, Mdn = 13). 4 studies reported participants ($< 0.02\%$ overall studies) that self-identified beyond female or male. We cannot provide any information on the other studies, where other genders could either not have been selectable options on the questionnaire or not represented in the sample at all. Furthermore, in cases where the studies specified any population that their sample was drawn from, it was a university context, i.e., students, staff, or other institutional members, in 56 of the cases (83%).

4.5 Discussion

Our literature review of 61 papers presenting experiments with artificial locomotion techniques revealed several deficiencies impacting their reproducibility. In terms of the **experimental setup**, a large proportion of papers lacked details on the exact implementation of their locomotion techniques (67% for steering, 89% for teleportation). While the descriptions of participant samples were often sufficient to gain an initial impression, we observed large differences in the reported measurements of age and prior VR experience, which complicates comparisons across publications. Regarding the **scientific rigor**, the majority of papers did not specify their research question explicitly (69%), and a considerable part applied NHST without the specification of hypotheses (39%). Furthermore, the reports of statistical results were incomplete in approximately half of the cases (51%). Concerning **generalizability**, while the average sample size of 25.1 was higher than reported in related disciplines, the samples were often biased by overrepresenting young, male, and academic users.

As a consequence of these issues, we believe that a reproduction project as done in other disciplines (e.g., [27]) would often find it difficult to (1) reconstruct the original setup of an experiment accurately and (2) generate the same experimental results and conclusions even if the setup can be successfully reconstructed. We are concerned by these findings and would like to increase the awareness of this issue within the VR locomotion community. However, it is important to mention that our goal of this work is not to pick on individual publications for their limitations but to identify more overarching issues in the existing practices pursued in the VR locomotion research community. To strengthen this point, we would like to emphasize that the presented survey also included our own publications, which also did not receive positive ratings on all of the surveyed aspects.

We currently have two possible explanations for why the results look so surprisingly alarming, especially given that we have only scratched the surface by looking at a few selected attributes. First, we believe that the abstraction of the underlying concepts (here steering and teleportation) necessary for complete reporting and the associated step backward to classify one's own work in the existing corpus of other works often do not take place sufficiently. From our own experience, however, we also know that such an abstraction is not easy under the constant change of technology. For example, it is often not considered worth mentioning whether user rotations during navigation are performed physically or virtually. While physical rotation was deemed the standard in times when CAVEs (with 360° field of regard) were among the dominant projection devices for VR, virtual rotation is now considerably more prominent with HMDs [42].

This example introduces our second explanation. We suspect that a large part of the gaps is because information is simply considered obvious or that a non-existent standard is assumed. For example, one could assume that the majority of steering interfaces that only have a maximum speed specified probably interpolate linearly up to this speed. However, the reader cannot be sure as many other possibilities may even make more sense in certain use cases.

In an attempt to mitigate these issues in future VR locomotion experiments, we summarized the main insights of our literature review in the form of initial reproducibility guidelines, which are presented in the next section. While guidelines are a promising initial step towards improving the chances of successful reproduction, we acknowledge that identifying and providing every single detail relevant to the experimental outcome is highly challenging if not impossible. It is therefore always advisable to contact the original authors to clarify any ambiguities that arose while reading their work. Nevertheless, based on our experiences of conducting the literature review, we believe that following our guidelines can add clarity by providing readers with the most relevant information in clearly marked locations and a structured form of presentation. This provides the reader with more efficient information access compared to the process of scanning multiple and potentially long prosaic text passages.

5 REPRODUCIBILITY GUIDELINES

To support the communal effort towards increasing the reproduction chances of our experiments, we formulated guidelines to serve as a basis for future locomotion experiments.

5.1 Description of Experimental Setup

E_1: Provide an accurate description of the employed hardware setup, including the field-of-view, framerate, and motion-to-photon latency of the selected VR device(s). These factors exert substantial influence on experimental outcomes, particularly within the domain of VR locomotion research. Although our work of this paper primarily emphasizes the communication of software-related metrics and data analysis, the articulation of hardware attributes holds equivalent significance.

E_2: Align with existing abstractions and their naming conventions. In Sections 5.1.1 and 5.1.2 below, we provide basic abstractions for steering and teleportation interfaces.

5.1.1 Steering. For a steering interface to be reproducible, the reporting of a *direction specification* and a *speed specification* that manipulate the *viewer's position* are required.

Direction Specification. Often the direction $\in \mathbb{R}^3$ is directly derived from a tracked device's pose (input). In that case, it is important to specify these components precisely. For example, most of the *gaze-directed* interfaces do not track gaze or the user's head but the pose of the worn HMD, which can make a difference and cause confusion. Furthermore, all modifications of the input (*transfer function*) need to be specified. For example, 3D directional input often is projected onto surfaces to create, for instance, 2D ground-based steering interfaces. Directions are also derived from analog sticks, which are mounted on tracked devices. In these cases, it is important whether or not these directions are treated as global or local, with respect to the viewpoint's coordinate system.

Speed Specification. The specification of speed $\in \mathbb{R}$ does neither need to be continuous nor actively performed by the user. These cases should be explicitly stated. In all other cases, it is crucial to specify which input is used, how it looks like (e.g., binary $\in \{0, 1\}$, 1D axis $\in [0, 1]$, 1D axis $\in [-1, 1]$, body inclination in $^\circ$ or position offset $\in \mathbb{R}^3$, etc.), how it translates to the movement speed, and whether this happens, for instance, acceleration-based or by instantaneous/direct specification. This can be realized by the specification of a maximum speed **and** a transfer function. This transfer function does not need to be continuous or have a closed form, as is often the case with deadzones. In such situations, the piecewise definitions of the transfer function should be provided.

Rotation. In addition to manipulating the position, steering interfaces often also manipulate the *viewer's orientation.* If this is not the case, because the user is purely physically rotating, it should be explicitly stated to avoid ambiguity. Otherwise, the exact details of rotation speed and direction should be specified once again, usually in the form of a rotation axis and angular velocity along with their relation to the input. This reporting requirement is still true for situations where parameters are assumed to be "obvious", like the choice of the yaw axis for rotations in ground-based locomotion interfaces.

Travel Information. Finally, any visual or multi-modal feedback as well as other modifications, such as velocity-based field-of-view modifications [16], should be specified as precisely as possible.

5.1.2 Teleportation. The full report of a teleportation technique contains detailed information on the four mechanisms chosen for *target specification*, *pre-travel information*, *transition*, and *post-travel feedback* [40]. While the methods for target specification and transition form the core of each teleportation technique and should be reported in any case, additional pre-travel information and post-travel feedback are optional and therefore often not implemented. For these cases, we recommend mentioning the lack of these mechanisms explicitly to provide clarity to readers. If they are provided, they should be described as precisely as possible, but the exact details to report are largely dependent on the exact mechanisms that were developed. For the remaining phases, more general advice can be provided as follows:

Target Specification. The report of a target specification mechanism should explicitly mention the shape of the selection tool (e.g., straight ray, parabola) and the object that the tool was attached to (e.g., controller, head). If the shape of the selection tool is not a straight ray, additional parameters that describe the appearance of the tool as well as its maximum reach in the virtual environment should be provided. Furthermore, researchers should mention details about the selectable positions in the scene and the parts that were deliberately excluded (e.g., vertical surfaces).

Transition. The employed transition mode should be clearly mentioned by name (e.g., instantaneous, fade-to-black, animated). For other mechanisms than an instantaneous transition, additional details and parameters should be reported. The report of a fade-to-black transition, for example, should explicitly mention its duration while the report of an animated transition requires information on the type of animation (e.g., linear, slow-in-slow-out) as well as its maximum speed or total duration.

E_3: Actively consider what additional information would be necessary to re-implement the interface as abstractions can not capture the complete complexity of every interface.

E_4: Publish the source code alongside the paper, whenever possible. However, this requires significant additional effort, especially when it comes to clarifying issues around the co-publication of external resources. It is helpful to consider publishing the source code and all other necessary sources from the very beginning and thus consider the selection of external sources or the creation of own assets while the code is written. It is also important to acknowledge this additional effort from the reviewers' perspective.

E_5: Provide figures, drawings, or additional video material as they can help to solve ambiguities. However, it is important that these supplementary materials serve as a support of the written paper and not as a necessity for its understanding.

E_6: Describe the population that your participants were recruited from and specify the sampling strategy that was used. This can be *convenience sampling*, when participants are recruited who are readily available and accessible, up to *stratified sampling* which involves dividing the target population into homogeneous subgroups based on specific characteristics like age, gender, or other relevant characteristics.

E_7: Describe your sample by providing at least a gender, age, and VR experience distribution. The following two guidelines give explicit advice on how this can look for age and prior VR experience.

E_8: Use the mean, median, standard deviation, and range to describe your sample's age distribution. Different instruments have different advantages and disadvantages, such as the susceptibility of the mean to be influenced by extreme values, which complement each other. Providing multiple instruments also helps to increase compatibility with other research papers.

E_9: Report multiple variables to describe participants' prior experience with VR. A boolean classifier (i.e., the specification if participants had prior VR experience or not) is a good first step because it is widely used. In order to increase the informative value, however, further information should be provided. The large number of different measures we have seen in our literature review shows that quantifying prior experience is not straightforward. Again, the only remedy is to raise and report different measures in parallel. We suggest using a combination of three measures that together represent prior exposure to VR [19]: frequency (e.g., *never*, *daily*, *once a week*, etc.), tenure (e.g., *never*, *less than a year*, *1-2 years*, etc.), and depth/diversity of use (e.g., *Which of the following 3D user interfaces have you already used?*). As an alternative or an additional point, one can inquire about self-perceived competence [19]. Recent work by Steed et al. also suggested to devise novel measurements that distinguish immersive *competence* from immersive *literacy* as two related but separate measures of prior VR experience [37].

E_{10}: Report the balancing strategy and a separate sample description for each individual group in between-subject designs. This helps the reader to understand potential confounding variables introduced by inter-group differences, especially when the overall sample size is small.

5.2 Scientific Rigor of the Research Process

R_1: State the research question(s) of your work explicitly in the introduction. While several papers argue about the rationale of their work in their introduction, the explicit mention of the research question(s) adds clarity to help readers understand the focus of the research. It also helps with validating the chosen experimental design, which is not possible when the research question(s) has or have to be derived from the experimental design itself.

R_2: Formulate hypotheses before the experiment and use them to structure the statistical evaluation. The pre-formulation of hypotheses is a central requirement of the scientific process, and registering hypotheses on online platforms can provide additional transparency as well as proof of your rigorous work. The structure of the paper's evaluation section should then follow the list of hypotheses and focus specifically on their validation. Additional statistical evaluations of non-hypothesized effects can be provided, but they should be labeled as exploratory insights to be confirmed by further studies.

R_3: Select hypotheses carefully and provide justifications for them. Particularly in the context of NHST, every significance test comes with a certain error probability. As a result, the number of hypotheses should be restricted to a set of well-justified statements rather than speculation of all potential differences in the data. The selected hypotheses should also be related to the research question(s). Furthermore, a low prior probability of effect (< 0.75) can quickly cause a proportion of false positive results massively exceeding the rate of 5% [28], which initially was aimed for by choosing an alpha level of 0.05.

R_4: Consider preregistration of the experimental design on online platforms like OSF. This helps to prove, for instance, that the hypotheses were formulated prior to the experiment. While this process has also been criticized for impairing exploratory analyses beyond the registered hypotheses [36], the general consensus is that exploratory research may still be conducted and reported if clearly labeled as such [31].

R_5: Provide a full report of your NHST results. This includes the name of the test statistic (e.g., t, F), the degrees of freedom of the test, the value of the test statistic, and the p-value. This is also true for non-significant results as it gives readers insights into the exact figures behind your conclusions and enables them to conduct future meta and power analyses for their own work.

R_6: Consider alternative metrics to complement or replace NHST. The report of effect sizes and confidence intervals together with a graphical depiction of your data can provide more insights than a dichotomous significance test. Effect sizes are also helpful to assist future researchers to acquire an estimated effect size for their power analysis.

5.3 Generalizability of Results

G_1: Consider diverse user groups in your user studies. The ideal sample is large and has an equal distribution of gender, age, prior VR experience, and ethnicity.

G_2: Be upfront about the potential limitations of your conclusions. Gathering a representative sample of the general population is a challenging process. Discussing the limitations of your conclusions and detailing the restricted subset of the general population that they are likely to apply to, helps to put your results into context and simplifies the identification of potential areas for future work.

G_3: Be open to supporting other researchers attempting a reproduction of your work. Beyond your attempts of writing about your techniques and experiments in a reproducible form, be willing to share more details and answer questions when contacted by other researchers.

6 LIMITATIONS

While our work presented in this paper started with the identification of general replication challenges for VR locomotion research, our literature review was limited to artificial locomotion techniques presented at three conferences in the last five years. As a result, our literature review cannot be generalized to physical locomotion techniques (i.e., the second type in the taxonomy of Boletsis and Chasanidou [4]) like walking-in-place and redirected walking. The restriction of venues was necessary to keep the size of the resulting paper collection manageable and to make comparisons between papers within the collection easier. From our perspective, we have selected the three most influential conferences focused on VR technologies. Another or extended set of conferences or an even stronger focus on journals could yield different results, but we see no reason to assume that these results would substantially differ to an extent that would negate the observations we have made. The limitation of the time frame is justified by our research question, which inquires about the reproducibility of the current state-of-the-art. Equally interesting would be questions that consider longer time spans or even attempt to identify trends. However, these cannot be addressed with our subsample here.

Another limitation comes from the fact that the papers in our literature review were all selected, annotated, and analyzed manually, which does not rule out potential influences of human error. While we took great care to be as rigorous as possible, extracting all relevant information from the papers was often challenging due to the lack of standardized reporting and sectioning, resulting in a leftover risk of having overseen certain pieces of information in longer text paragraphs. Furthermore, the transfer of information into comparable quantifiable attributes was also based on human judgment and therefore might be disputable in individual cases. However, given the clear picture painted by our results, we do not believe that these and similar minor lapses due to human error would have an influence on our overall conclusions.

Moreover, we are aware that we restricted our analysis to only a restricted set of factors influencing reproducibility. The successful reproduction of experiments is a complex process involving various known and unknown variables even within a certain discipline. Therefore, our formulated guidelines are only an incomplete starting point to improve the reproducibility of artificial locomotion experiments, and we invite other researchers to contribute to this list in the future.

Finally, while we formulated initial guidelines on how the reproducibility of experiments can be improved, we recognize that following selected guidelines might be challenging under certain circumstances. For example, capturing demographic data can be restricted by certain ethics committees if it does not contribute to the primary focus of the study.

7 CONCLUSION AND FUTURE WORK

We presented reproducibility challenges for empirical experiments in VR locomotion research by highlighting the importance of the description of the experimental setup, the scientific rigor of the research process, and the generalizability of results. Based on our formal literature review of artificial locomotion experiments published at three major international venues in the last five years, we conclude that state-of-the-art research does often not fulfill basic requirements for reproducibility. As a result, we as the VR locomotion research community should stay vigilant when reading and writing about our empirical experiments. We believe that our guidelines provided in this paper serve as a solid first step towards improving current practices, and we invite other researchers to comment on and extend these guidelines in the future. As a community, we have the collective power to counteract a replication crisis and thus make our research findings more impactful.

REFERENCES

[1] Majed Al Zayer, Paul MacNeilage, and Eelke Folmer. 2020. Virtual Locomotion: A Survey. *IEEE Transactions on Visualization and Computer Graphics* 26, 6 (2020), 2315–2334. https://doi.org/10.1109/TVCG.2018.2887379

[2] Louise Barkhuus and Jennifer A. Rode. 2007. From Mice to Men - 24 Years of Evaluation in CHI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/1240624.2180963

[3] Lonni Besançon and Pierre Dragicevic. 2019. The Continued Prevalence of Dichotomous Inferences at CHI. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–11. https://doi.org/10.1145/3290607.3310432

[4] Costas Boletsis and Dimitra Chasanidou. 2022. A Typology of Virtual Reality Locomotion Techniques. *Multimodal Technologies and Interaction* 6, 9 (2022), 14 pages. https://doi.org/10.3390/mti6090072

[5] Doug A. Bowman, David Koller, and Larry F. Hodges. 1997. Travel in Immersive Virtual Environments: An Evaluation of Viewpoint Motion Control Techniques. In *Proc. of IEEE VR* (Albuquerque, NM, USA). IEEE, New York, NY, USA, 45–52.

[6] Doug A. Bowman, Ernst Kruijff, Joseph J. LaViola, and Ivan Poupyrev. 2001. An Introduction to 3-D User Interface Design. *Presence* 10, 1 (2001), 96–108. https://doi.org/10.1162/105474601750182342

[7] Evren Bozgeyikli, Andrew Raij, Srinivas Katkoori, and Rajiv Dubey. 2016. Point & Teleport Locomotion Technique for Virtual Reality. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play* (Austin, Texas, USA). ACM, New York, NY, USA, 205–216. https://doi.org/10.1145/2967934.2968105

[8] Hugo Brument, Gerd Bruder, Maud Marchal, Anne Hélène Olivier, and Ferran Argelaguet. 2021. Understanding, Modeling and Simulating Unintended Positional Drift during Repetitive Steering Navigation Tasks in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics* 27, 11 (2021), 4300–4310. https://doi.org/10.1109/TVCG.2021.3106504

[9] Kelly Caine. 2016. Local Standards for Sample Size at CHI. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 981–992. https://doi.org/10.1145/2858036.2858498

[10] Lorys Castelli, Luca Latini Corazzini, and Giuliano Carlo Geminiani. 2008. Spatial navigation in large-scale virtual environments: Gender differences in survey tasks. *Computers in Human Behavior* 24, 4 (2008), 1643–1667. https://doi.org/10.1016/j.chb.2007.06.005

[11] Andy Cockburn, Pierre Dragicevic, Lonni Besançon, and Carl Gutwin. 2020. Threats of a Replication Crisis in Empirical Computer Science. *Commun. ACM* 63, 8 (2020), 70–79. https://doi.org/10.1145/3360311

[12] Simon Davis, Keith Nesbitt, and Eugene Nalivaiko. 2014. A Systematic Review of Cybersickness. In *Proceedings of the 2014 Conference on Interactive Entertainment* (Newcastle, NSW, Australia) *(IE2014)*. Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/2677758.2677780

[13] Sarah Delgado Rodriguez, Radiah Rivu, Ville Mäkelä, and Florian Alt. 2023. Challenges in Virtual Reality Studies: Ethics and Internal and External Validity. In *Proceedings of the Augmented Humans International Conference 2023*. ACM, New York, NY, USA, 105–111. https://doi.org/10.1145/3582700.3582716

[14] Randi A. Doyle, Daniel Voyer, and Isabelle D. Cherney. 2012. The relation between childhood spatial activities and spatial abilities in adulthood. *Journal of Applied Developmental Psychology* 33, 2 (2012), 112–120. https://doi.org/10.1016/j.appdev.2012.01.002

[15] David Englmeier, Wanja Sajko, and Andreas Butz. 2021. Spherical World in Miniature: Exploring the Tiny Planets Metaphor for Discrete Locomotion in Virtual Reality. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. IEEE, New York, NY, USA, 345–352. https://doi.org/10.1109/VR50410.2021.00057

[16] Ajoy S. Fernandes and Steven K. Feiner. 2016. Combating VR Sickness Through Subtle Dynamic Field-Of-View Modification. In *2016 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, New York, NY, USA, 201–210. https://doi.org/10.1109/3DUI.2016.7460053

[17] Andy Field. 2013. *Discovering Statistics Using IBM SPSS Statistics*. SAGE Publications, Newbury Park, California, USA.

[18] Kasper Hornbæk, Søren S. Sander, Javier Andrés Bargas-Avila, and Jakob Grue Simonsen. 2014. Is Once Enough? On the Extent and Content of Replications in Human-Computer Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 3523–3532. https://doi.org/10.1145/2556288.2557004

[19] Jörn Hurtienne, Anne-Marie Horn, and Pat Langdon. 2010. Facets of Prior Experience and Their Impact on Product Usability for Older Users. In *Designing Inclusive Interactions: Inclusive Interactions Between People and Products in Their Contexts of Use*. Springer, London, UK, 123–132. https://doi.org/10.1007/978-1-84996-166-0_12

[20] Wonil Hwang and Gavriel Salvendy. 2010. Number of People Required for Usability Evaluation: The 10±2 Rule. *Commun. ACM* 53, 5 (2010), 130–133. https://doi.org/10.1145/1735223.1735255

[21] Maurits Kaptein and Judy Robertson. 2012. Rethinking Statistical Analysis Methods for CHI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1105–1114. https://doi.org/10.1145/2207676.2208557

[22] Norbert L Kerr. 1998. HARKing: Hypothesizing After the Results are Known. *Personality and Social Psychology Review* 2, 3 (1998), 196–217. https://doi.org/10.1207/s15327957pspr0203_4

[23] Joseph J. LaViola Jr, Ernst Kruijff, Ryan P. McMahan, Doug Bowman, and Ivan P. Poupyrev. 2017. *3D User Interfaces: Theory and Practice*. Addison-Wesley Professional, Boston, Massachusetts, USA.

[24] Maud Marchal, Julien Pettré, and Anatole Lécuyer. 2011. Joyman: A Human-Scale Joystick for Navigating in Virtual Worlds. In *2011 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, New York, NY, USA, 19–26. https://doi.org/10.1109/3DUI.2011.5759212

[25] Thinh Nguyen-Vo, Bernhard E. Riecke, Wolfgang Stuerzlinger, Duc-Minh Pham, and Ernst Kruijff. 2021. NaviBoard and NaviChair: Limited Translation Combined with Full Rotation for Efficient Virtual Locomotion. *IEEE Transactions on Visualization and Computer Graphics* 27, 1 (2021), 165–177. https://doi.org/10.1109/TVCG.2019.2935730

[26] Anna Offenwanger, Alan John Milligan, Minsuk Chang, Julia Bullard, and Dongwook Yoon. 2021. Diagnosing Bias in the Gender Representation of HCI Research Participants: How It Happens and Where We Are. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 399, 18 pages. https://doi.org/10.1145/3411764.3445383
[27] Open Science Collaboration. 2015. Estimating the Reproducibility of Psychological Science. *Science* 349, 6251 (2015), aac4716. https://doi.org/10.1126/science.aac4716
[28] Harold Pashler and Christine R. Harris. 2012. Is the Replicability Crisis Overblown? Three Arguments Examined. *Perspectives on Psychological Science* 7, 6 (2012), 531–536. https://doi.org/10.1177/1745691612463401
[29] Tabitha C. Peck, Laura E. Sockol, and Sarah M. Hancock. 2020. Mind the Gap: The Underrepresentation of Female Participants and Authors in Virtual Reality Research. *IEEE Transactions on Visualization and Computer Graphics* 26, 5 (2020), 1945–1954. https://doi.org/10.1109/TVCG.2020.2973498
[30] Aniruddha Prithul, Isayas Berhe Adhanom, and Eelke Folmer. 2021. Teleportation in Virtual Reality; A Mini-Review. *Frontiers in Virtual Reality* 2 (2021), 7 pages. https://doi.org/10.3389/frvir.2021.730792
[31] Patrick Langford (PsychBrief). 2016. In Defence of Preregistration. https://psychbrief.com/2016/05/24/in-defense-of-preregistration [last accessed 2023-08-11].
[32] Kasra Rahimi, Colin Banigan, and Eric D. Ragan. 2020. Scene Transitions and Teleportation in Virtual Reality and the Implications for Spatial Awareness and Sickness. *IEEE Transactions on Visualization and Computer Graphics* 26, 6 (2020), 2273–2287. https://doi.org/10.1109/TVCG.2018.2884468
[33] Lisa Rebenitsch and Charles Owen. 2014. Individual Variation in Susceptibility to Cybersickness. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* (Honolulu, Hawaii, USA) *(UIST '14)*. Association for Computing Machinery, New York, NY, USA, 309–317. https://doi.org/10.1145/2642918.2647394
[34] Robert Rosenthal. 1979. The File Drawer Problem and Tolerance for Null Results. *Psychological Bulletin* 86, 3 (1979), 638. https://doi.org/10.1037/0033-2909.86.3.638
[35] Martin Schmettow. 2012. Sample Size in Usability Studies. *Commun. ACM* 55, 4 (2012), 64–70. https://doi.org/10.1145/2133806.2133824
[36] Mel Slater. 2016. Pre-Registration of Analysis of Experiments is Dangerous for Science. https://presence-thoughts.blogspot.com/2016/04/pre-registration-of-analysis-of.html [last accessed 2023-08-11].
[37] Anthony Steed, Dan Archer, Lisa Izzouzi, Nels Numan, Kalila Shapiro, David Swapp, Dinah Lammiman, and Robert W Lindeman. 2023. Immersive Competence and Immersive Literacy: Exploring How Users Learn About Immersive Experiences. *Frontiers in Virtual Reality* 4 (2023). https://doi.org/10.3389/frvir.2023.1129242
[38] Cheryl Techentin, Daniel Voyer, and Susan D. Voyer. 2014. Spatial Abilities and Aging: A Meta-Analysis. *Experimental Aging Research* 40, 4 (2014), 395–425. https://doi.org/10.1080/0361073X.2014.926773
[39] Jeffrey C. Valentine, Ariel M. Aloe, and Timothy S. Lau. 2015. Life After NHST: How to Describe Your Data Without "p-ing" Everywhere. *Basic and Applied Social Psychology* 37, 5 (2015), 260–273. https://doi.org/10.1080/01973533.2015.1060240
[40] Tim Weissker, Andre Kunert, Bernd Froehlich, and Alexander Kulik. 2018. Spatial Updating and Simulator Sickness During Steering and Jumping in Immersive Virtual Environments. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, New York, NY, USA, 97–104. https://doi.org/10.1109/VR.2018.8446620
[41] Daniel Zielasko. 2021. Subject 001 - A Detailed Self-Report of Virtual Reality Induced Sickness. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, New York, NY, USA, 165–168. https://doi.org/10.1109/VRW52623.2021.00038
[42] Daniel Zielasko, Jonas Heib, and Benjamin Weyers. 2022. Systematic Design Space Exploration of Discrete Virtual Rotations in VR. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, New York, NY, USA, 693–702. https://doi.org/10.1109/VR51125.2022.00090
[43] Daniel Zielasko, Sven Horn, Sebastian Freitag, Benjamin Weyers, and Torsten W. Kuhlen. 2016. Evaluation of Hands-Free HMD-Based Navigation Techniques for Immersive Data Analysis. In *Proc. of IEEE Symposium on 3D User Interfaces*. IEEE, New York, NY, USA, 113–119. https://doi.org/10.1109/3DUI.2016.7460040
[44] Daniel Zielasko, Yuen C. Law, and Benjamin Weyers. 2020. Take a Look Around – The Impact of Decoupling Gaze and Travel-direction in Seated and Ground-based Virtual Reality Utilizing Torso-directed Steering. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, New York, NY, USA, 398–406. https://doi.org/10.1109/VR46266.2020.00060