You're Making Me Sick

A Systematic Review of How Virtual Reality Research Considers Gender & Cybersickness

Cayley MacArthur
University of Waterloo
Waterloo, Ontario, Canada
cayley.macarthur@uwaterloo.ca

Arielle Grinberg
University of Waterloo
Waterloo, Ontario, Canada
arielle.grinberg@uwaterloo.ca

Daniel Harley
University of Waterloo
Waterloo, Ontario, Canada
dharley@uwaterloo.ca

Mark Hancock
University of Waterloo
Waterloo, Ontario, Canada
mark.hancock@uwaterloo.ca

ABSTRACT

While multiple studies suggest that female-identified participants are more likely to experience cybersickness in virtual reality (VR), our systematic review of 71 eligible VR publications (59 studies and 12 surveys) pertaining to gender and cybersickness reveals a number of confounding factors in study design (e.g., a variety of technical specifications, tasks, content), a lack of demographic data, and a bias in participant recruitment. Our review shows an ongoing need within VR research to more consistently include and report on women's experiences in VR to better understand the gendered possibility of cybersickness. Based on the gaps identified in our systematic review, we contribute study design recommendations for future work, arguing that gender considerations are necessary at every stage of VR study design, even when the study is not 'about' gender.

CCS CONCEPTS

• Human-centered computing → HCI design and evaluation methods; **Virtual reality**; **• Social and professional topics → Gender**.

KEYWORDS

gender, sex, cybersickness, simulator sickness, virtual reality, virtual environments, systematic review

ACM Reference Format:
Cayley MacArthur, Arielle Grinberg, Daniel Harley, and Mark Hancock. 2021. You're Making Me Sick: A Systematic Review of How Virtual Reality Research Considers Gender & Cybersickness. In *CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3411764.3445701

1 INTRODUCTION

For decades, research into virtual reality (VR) has provided evidence that women may be disproportionately affected by the negative symptoms of cybersickness [47, 78, 81], ranging from discomfort to the possibility of an emetic response. Research into cybersickness follows a longer history into simulator sickness[1], which also provides preliminary evidence that women may be more susceptible to the possible negative effects of VR. Despite this long history, cybersickness remains an unsolved problem for VR, and the apparent link between cybersickness and women across VR research has not been systematically studied or validated. With the recent release of relatively low-cost consumer head-mounted displays (HMDs), there is an increased availability and access to VR technologies, both for the public and for the research community. VR technologies and applications impact a wide variety of domains [23, 54], with early evidence suggesting a range of prosocial benefits [28, 68]. Yet the possibility of negative symptoms remains a consistent and often overlooked concern. Given the persistent unknowns about the safety or risks associated with VR, it is crucial that more comprehensive and/or targeted study design methods and practices are developed to better understand the relationship between gender and cybersickness. With an initial focus on human-computer interaction (HCI), we propose to examine how VR research accounts for the possible gendered effects of cybersickness. We then ask how we might leverage lessons learned from these publications in order to better understand the possible relationship between VR, gender, and cybersickness. If VR is to live up to its promise as a technology for 'everyone' [65, 88], it is integral to find ways to ensure that it does not exclude by design.

We present a systematic review of literature that comments on the connections between gender, cybersickness, and VR in order to propose equitable study design guidelines for VR. Our systematic review is informed by our preliminary review of CHI 2019 VR papers that shows that considerations for participant gender are inconsistent and underreported. Expanding this search, our systematic review of 71 eligible VR publications (59 studies and 12 surveys) draws on literature across a variety of related fields (health, human factors, psychology, and so on) in order to derive more generalizable recommendations for study design and to better understand the

CHI '21, May 8–13, 2021, Yokohama, Japan
The definitive Version of Record was published in *CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan*, https://doi.org/10.1145/3411764.3445701.

---

[1] The terms *simulator sickness* and *cybersickness* were often used interchangeably over the course of our review, but they are not the same: *cybersickness* pertains to the discomfort felt during or after using a virtual environment (VE), with approximately three times the severity and a different symptom profile to *simulator sickness*, which pertains to simulator environments [47, 84]. We also encountered attributions of sickness to both *gender* and *sex* in our review. Given that the standard manner of establishing participant gender is by self-report, we use *gender* in our work to reflect and respect the identity disclosed by participants.

relationship between gender and cybersickness. It should be noted that our systematic review is unlike other traditional systematic reviews. Our findings reveal that most VR studies within HCI do not report on gender effects, so we were unable to provide statistical analyses on our topics of interest. This does, however, show the need within the VR community to acknowledge the gendered effects on cybersickness. We argue that with the possibility of a gendered susceptibility to cybersickness, the VR community must consider gender at the forefront of study design. Our work reveals a number of confounding factors (e.g., a wide variety of technical specifications, tasks, content), a lack of demographic data, and a bias in participant recruitment, which can make it impossible to identify or ascertain the specific causes or effects of gendered cybersickness. Furthermore, we uncovered a clear bias in inclusion and exclusion criteria, with a lack of data on those who are not able to participate due to the negative effects of cybersickness. Overall, these results are then reinforced by survey papers that repeat these findings, and perpetuated by further research that is informed by these surveys.

In our recommendations, we argue that there is a need for more consistent study design and reporting. The principles of human-participant research suggest that representative samples of demographically diverse participants lead to more generalizable findings [46, 68]. Moreover, we suggest that there is a need for this research to adopt more nuanced perspectives of sex and gender, as socially-constructed and/or biological characteristics appear to be assumed or essentialized. While our review confirms that multiple studies claim that female-identified participants are more likely to experience cybersickness in VR, the inconsistencies and the lack of clear study design guidelines that acknowledge and/or address this gender imbalance suggests opportunities for future work. Based on the gaps identified in our systematic review, we contribute preliminary study design recommendations, arguing that gender considerations are necessary at every stage of VR study design, even when the study is not 'about' gender.

2 BACKGROUND

Prior to our CHI 2019 review (Section 3) and our in-depth systematic review (Section 4), we explored the problem space of gender issues in VR that inspired the work in our paper. In this informal literature review, we observed that publications coalesced into three broad categories:

1. publications *documenting and comparing* gender and susceptibility to sickness (e.g., [24, 66]);
2. research *investigating* the nature of gender differences that could lead to virtual reality (VR) discomfort, such as differences in path integration and visual dependency (e.g., [12, 26]); and
3. *interventions* that attempt to address suspected causes of the discomfort (e.g., [97]).

However, we also observed a misalignment between these three areas in terms of the measures used and the solutions proposed. For example, we found that (1) and (2) do not account for possible gender differences in symptom manifestation or expression (e.g., comparing "nausea" as one universal experience, as opposed to having multiple dimensions which differ in expression and prominence in the context of cybersickness [61, 83]), and (3) proposes universal solutions to cybersickness without validating their effectiveness among diverse groups. While still valuable, as a large amount of participants in VR research are likely to experience discomfort using VR systems [65, 88], creating universalizing solutions does not address the specific problems stated in (1) and (2). In this section, we expand on this misalignment by elaborating on categories 1-3, highlighting the issues we identified that inspired the systematic review we present in this paper.

In tandem with the paucity of research on this topic, and the difficulty of reconciling disparate findings across these categories, there is a lack of engagement with gender as a factor across study design and analysis that could negatively impact understandings of participant experiences in VR. We argue that this lack of consistency and/or attention to the possible gendered effects of cybersickness signals the need for future work. This, in turn, prompted our preliminary recommendations for future VR research.

2.1 Documenting and Comparing: Symptom Profiles Differ by Gender

Much of today's knowledge about gendered experiences of simulator sickness or cybersickness—the terms used to describe the combination of uncomfortable symptoms associated with virtual environment exposure—comes from the human factors field, where simulators were adopted for training in aerospace and defence. Stanney et al. [83] provide a comprehensive overview of what cybersickness—the term used to differentiate sickness from VR versus a simulator—often entails: dizziness, drowsiness, headache, nausea, fatigue, general malaise, and aftereffects including disturbed proprioception and postural instability. The most common theory for why cybersickness occurs is *sensory conflict theory* [70]. This theory positions cybersickness as a reflex (such as nausea) to a stimulus (conflicting sensory inputs): e.g., the visual system perceives motion while wearing a head-mounted display (HMD, a VR headset) while the body remains stationary.

In prior literature on gender and cybersickness, we found that symptom profiles are experienced differently by gender. For example, Stanney et al. [83] presents a study with 1102 participants that found that female participants experienced 15% higher total severity of sickness symptoms. Among other findings, they found that female and male participants experienced different "symptom profiles," that is, the symptoms recorded had a significantly different hierarchy of the experienced severity between sexes. In the order of most to least severe symptoms, women had a *Disorientation* > *Oculomotor Disturbances* > *Nausea* ($D > O > N$) symptom profile, while men had a *Disorientation* > *Nausea* > *Oculomotor Disturbances* ($D > N > O$) profile. In other words, women may experience less nausea than men during experiments while still experiencing overall more severe symptoms of cybersickness. Later research continues to find a higher level of total severity in women, but tends to overlook this possible difference in symptom profiles, resulting in the use of measures that do not fully capture what is happening between groups. Essentializing claims about cybersickness are likely to disregard the nuances of these profiles or differences in individual experiences. We contextualize this point with an example in Section 2.2.

2.2 Investigating Gender and Cybersickness: A Need to Document Women's Experiences of Discomfort

Some attempts have been made to distinguish whether women are simply more likely to report discomfort (due to reasons such as socialization or tolerance), rather than experiencing it more frequently. Jokerst et al. [34] attempted to rule out an effect of socialization, and in their study found that while the gender of the participant and the researcher did not significantly affect the participant's likelihood to report, female participants had higher symptom scores and reported significantly higher gastrointestinal symptoms in a post-survey. And yet, the authors doubted the validity of female participants' self-reports, with no mention of reasoning behind their doubts, and so they recorded gastric myoelectric activity to quantify levels of nausea. They found no difference between genders on the gastric measure. Similarly, Cheung and Hofer [14] use physiological measures (heart rate, blood pressure, etc.) and were unable to detect gender differences, and yet because the blood flow measures could not explain the reports, they state that women must be more inclined to report discomfort. Park and Hu [66] present a study of similar design, with a similar conclusion. In addition to the lack of data to support these explanatory efforts, this view of cybersickness neglects an account of the experience beyond nausea, especially given that nausea is not solely due to gastrointestinal distress.

Moreover, recall that in the female symptom profile, nausea was the least prominent symptom ($D > O > N$, [83]), and the sensation of nausea does not necessarily induce an emetic response, as would be detected by gastric sensors. "Nausea" is highly complex and contextual, with three distinct dimensions: somatic distress, gastrointestinal distress, and emotional distress [61]. According to Stanney et al. [83], while women do report more sickness than men overall, they do not experience more nausea than men, and differences between male and female participants were attributable to significantly higher levels of disorientation and oculomotor disturbances.

Further complicating reports of discomfort is the possibility of self-exclusion. For example, Flanagan et al. [24] control for a number of factors including willingness to volunteer given a history of motion sickness. The authors propose that a past history of motion sickness induces anxiety which exacerbates negative symptoms. They point out the numerous differences in questionnaires and lab studies in prior work, and cite evidence against the idea that men are more reticent to report motion sickness. Flanagan et al. [24] also suggest that there is a fundamental flaw in any research involving people who may be susceptible to motion sickness as these people might self-exclude, raising questions about how to better include those who may be most at risk. Taken together, these cases suggest a need to better document gendered experiences of discomfort.

2.3 Interventions: A Need for More Attention to Gender

While attempts have been made to mitigate cybersickness in VR, there is a lack of attention to the possibility of gendered effects of cybersickness across study design and analysis. Current research to reduce cybersickness includes examinations of walking in VR [98], or examinations of the effects of vertical axis alignment in supine postural VR use [90]. Yet without specific attention to gender, it is difficult to assess whether such efforts can be generalized, or how findings may relate to the possibility that women are more likely to experience discomfort in VR [83].

Comparing past research findings to contemporary experiences with VR may be necessary to provide more insight into the role that gender plays in these contexts. For example, one opportunity for closer examination is field of view (FOV) in VR. In the early 2000s, research showed that the size of display improved people's path integration ability in 3D virtual navigation tasks [89]. The authors hypothesized that the induced immersion caused by a larger display would influence participants to use "more efficient" egocentric navigation strategies. Similarly, Czerwinski et al. [17] argued that a wider field of view coupled with larger displays for navigating 3D virtual worlds improved women's navigation speed performance (note, however, that these were not head-mounted displays). Despite these apparent advantages for navigation, other research showed that a virtual reality display with a wide FOV can induce cybersickness more easily than a display with a narrow FOV [45, 49]. More recent research complicates these latter findings. Xiao and Benko [97] report that the relatively low-cost addition of sparse peripheral displays to existing headsets expands FOV with the unexpected result of reducing nausea symptoms. In this case, only 6 of the 17 participants were female, and gender was not part of the analysis. While this case benefits from putting the work into conversation with past research into FOV and cybersickness, the same is needed with regard to FOV and gender. For example, Al Zayer et al. [1] report that restricting the FOV is an effective mitigation strategy for cybersickness among both male and female participants.

While recent efforts to mitigate cybersickness are needed, the lack of direct attention to gender in such cases is a missed opportunity. Recent research, for example, argues that gender differences in cybersickness may be due to default interpupillary distance (IPD) in headsets, which is less likely on average to fit women compared to men [81]. If mitigation strategies are to be applied generally, there may be a need to consider how such approaches may have to change for diverse bodies and diverse individual experiences.

2.4 HCI Research into Gender Issues in VR

VR research in HCI commonly focuses on experiences in VR [31] and the usability of VR systems [88]. As such, VR studies in HCI that include a focus on gender appear to primarily describe experiences *within* VR: for example, gender swapping and avatar use [71, 77], experiences of harassment in VR [8, 62], or exploring sexuality and heteronormativity via pornography in VR [96].

Research on female representation within HCI has focused on the lack of female participants and authors in VR research papers [68], and the implications this might have on the field. These implications are vast, and authors note the importance of reporting data on gender and other demographics in order to better understand how each are affected by VR. The authors write, "Demographic information must be included when reporting characteristics of participants, including age, gender, and race/ethnicity, so that readers can accurately interpret the studied population and future meta-analyses of participant demographics can be performed" [68, p. 1952]. This recommendation has direct implications for systematic reviews such as this one, where inconsistent or missing participant data hinders an ability to perform such meta-analyses.

Research on usability stresses the importance of broad usability for VR to be applicable in educational settings, entertainment, job training, and more [23, 54]. However, while usability in VR often focuses on individual experiences, gender does not appear to be a key factor. Despite the significant efforts made by researchers to generate design recommendations to improve user experience and the usability of VR systems, few papers consider how and why gender may impact user experience.

Although gender *in relation* to cybersickness is not frequently investigated, discussions of cybersickness and the use of measures like the SSQ appear to be common in VR studies. Within HCI, research on experiences of cybersickness also includes exploratory work that provides guidelines and suggestions on design improvements for VR [20]. Again, there appears to be a lack of research into how and why gender may impact user experience.

After examining the research on gender within HCI, we were able to further identify the specific questions and gaps within the field. We specifically wanted to focus on how the CHI community, a well-known resource for VR research, reported cybersickness and the level of discomfort experienced in VR. This prompted our exploratory review of CHI 2019 papers, which we used to further develop the questions that we asked during our systematic review.

3 EXPLORATORY REVIEW

We began with an exploratory review of CHI 2019 papers to help develop a preliminary understanding of how current research addresses the possible relationship between VR, gender, and cybersickness. As the largest HCI conference, CHI provides a sample of current state-of-the-art research on VR. The overarching finding of this review is a lack of consistency with regard to how data on gender and/or cybersickness is collected and reported, which complicates more in-depth analysis across these studies. We describe the procedure we used to select papers and the insights that provided the foundation for the methods we applied to our systematic review (Section 4).

3.1 Search Procedure and Selection of Studies

Figure 1 shows the procedure of our review, from our exploratory review to the outcomes of our systematic review. Preceding our systematic review, we focused on CHI 2019 proceedings and papers from the ACM digital library. We used keywords such as "Virtual Reality", "Mixed Reality", and "VR" across paper titles and abstracts to identify eligible papers. We then refined our search to only include those that were directly about VR, excluding papers that were about investigating and/or testing a tool to be used with VR, such as controllers. Other papers that were excluded were those that primarily focused on other virtual environments, such as Augmented Reality. Two researchers analyzed the proceedings to ensure that only full papers were used (posters and demonstrations were excluded). The results yielded 96 papers.

3.2 Insights

Our analysis for the exploratory review focused on three overarching questions:

1. Did authors report the gender of their participants in terms of recruitment, and how many reported it in their results?
2. How was gender data collected?
3. Did authors measure, document, or report cybersickness and/or discomfort among participants?

In order to answer these questions, we looked at study classification types, sample size, gender of the participants, use of SSQ, contributions found about cybersickness, and whether the studies tested for differences in gender.

Overall, we found very little consistency across the 96 papers. For cybersickness, only 20 out of the 96 papers attempted to capture some aspect of participants' experience of sickness, and the strategies among these papers varied. 11 used the SSQ, while 9 used other measures: 1 used the Virtual-Reality Sickness Questionnaire (VRSQ), 1 used the Motion Sickness Assessment Questionnaire (MSAQ), 1 used questions from Witmer and Singer's presence questionnaire [95], and 6 used verbal feedback. Similarly, the papers presented a range of study types (e.g., user studies, pilot experiments, etc.), and were not consistent in the ways that they reported on gender. These inconsistencies suggest a need for further examination. For example, of the papers analyzed, only 65 reported participant gender, and only 3 reported their results across gender, meaning that it is not possible to identify specific issues or causes between gender and cybersickness within this recent work.

These inconsistencies led us to investigate the supplemental information of the papers in our exploratory review in order to determine the demographics of the participants, especially for studies that did not report on gender. Our goal was to assess whether we could use the supplemental information to pursue further gender-based analysis. However, of the 96 papers, only 2 provided supplemental information: 1 provided partial supplemental information, and 1 provided the questionnaires used. Broadly speaking, we were surprised not only by the inconsistencies, but also by how little information on gender and cybersickness was reported by authors within the community. Again, this suggests the need for further research to better understand a possible gendered susceptibility to cybersickness.

4 SYSTEMATIC REVIEW: METHOD

A common theme across our related work and exploratory review is the wide range of fields and the variety of possible applications that are impacted by VR research. Because the possibility of a gendered susceptibility to cybersickness is also a common theme, a strength of our approach is the ability to draw from findings that cross disciplines. To ensure that we reviewed a broad range of research, we used the following databases that have a history of publishing VR research: ACM DL, IEEE, PubMed, SagePub, as well as our institutional library databases which include SCOPUS and Web of Science. The search strings used in each database are presented in Table 1.

Using these strings returned a total of 662 records. At this first stage of screening per the PRISMA systematic review protocol [56], a group of 11 raters evaluated the records for eligibility based on

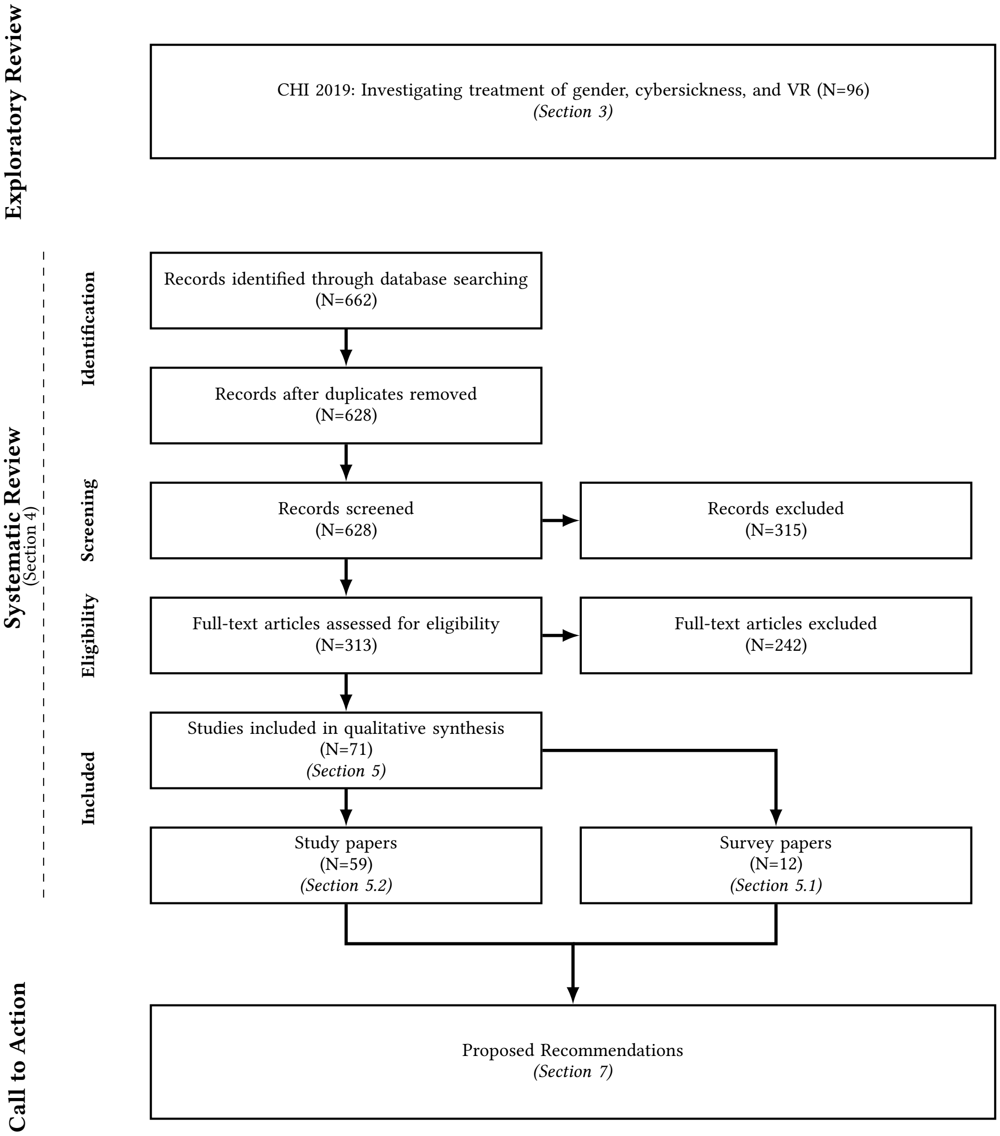

Figure 1: Number of results at each stage of the review, as represented in a PRISMA flow diagram.

titles and abstracts. Agreement was calculated at 98.8% between raters after independently rating 83 papers to ensure calibration before proceeding. In this first screening phase, publications were marked as follows:

- *Relevant*, wherein the title and/or abstract refer to the three main concepts of: gender or sex; cyber-, simulator-, or motion-sickness; and virtual or mixed reality;
- *Somewhat*, wherein the title and/or abstract refer to two of the three core concepts mentioned above;
- *Not Relevant*, wherein the title and/or abstract may include a reference to one of our core concepts or terms, but with no relation to the others, or the terms are being used in a different context.

As reflected in our PRISMA diagram (Figure 1), 315 papers were excluded at this screening stage (34 of which were duplicates). Of the remaining 313 papers:

- 66 papers met the criteria for *Relevant*, defined as "measures or finds something about the relationship between: gender or sex; cyber-, simulator-, or motion-sickness; and virtual or mixed reality," and were forwarded onwards for full-text analysis;
- 247 papers were deemed *Somewhat* relevant, defined as "needing further examination; may be about cybersickness or simulator sickness, but does not tell us about the sex or gender *relationship* to it; tells us about sex or gender and motion sickness or cybersickness (but is not VR/MR); or uses the key terms in passing but does not contribute new information,"

Database	Search string	Results	Date of search
ACM Digital Library	`[[All: "simulator sickness"] OR [All: "cybersickness"]] AND [[All: gender] OR [All: sex]] AND [[All: "virtual reality"] OR [All: "virtual environments"] OR [All: "mixed reality"]]`	191	February 19, 2020
IEEE Xplore Digital Library	`(gender OR sex) AND ("simulator sickness" OR "cybersickness") AND ("virtual reality" OR "virtual environments" OR "mixed reality")`	7	February 19, 2020
PubMed Central (PMC)	`(gender OR sex) AND ("simulator sickness" OR "cybersickness") AND ("virtual reality" OR "virtual environments" OR "mixed reality")`	155	February 12, 2020
SagePub Journals	`(gender OR sex) AND ("simulator sickness" OR "cybersickness") AND ("virtual reality" OR "virtual environments" OR "mixed reality")`	101	February 18, 2020
Institution Library	`(gender OR sex) AND ("simulator sickness" OR "cybersickness") AND ("virtual reality" OR "virtual environments" OR "mixed reality")`	208	February 19, 2020

Table 1: Databases reviewed and search strings used for the systematic review.

these required further investigation beyond title and abstract, and these 247 papers were evaluated based on full-text contents by a group of 11 raters (including 3 of the authors), all with previous familiarity with the project and its goals from the prior screening stage.

Following this full-text analysis, 242 papers were excluded for reasons including: the publication is not in English and a reliable translation was requested but not obtainable by the authors; the research is not about humans; the publication is not a full paper. Finally, we arrived at a list of 71 *Relevant* articles, comprising 59 studies and 12 survey papers.

4.1 Phase 1: Data Collection

After narrowing down the set of publications to be analyzed, we collectively defined a number of categories to log details from the papers for further analysis based on the studies reviewed. The exhaustive list of categories is as follows: display type; stereoscopic; motion tracking; driving simulator; commercial or custom; study design (between, within, or mixed participants); sample size; number and percentage of female participants; whether gender or sex were the primary focus of study or supplementary findings; whether any non-binary understanding of gender was expressed by the authors; if any gender or sex differences were found; number of dropouts and dropout gender; measures of sickness; independent and dependent variables; participant demographics reported; type of exposure; duration of exposure; exclusion criteria; content (when available); what relationship (if any) was found between gender and cybersickness. We also noted any overall strengths and limitations of the studies as they pertained to answering the questions guiding this systematic review:

1. How does VR research account for the possible gendered effects of cybersickness?
2. What can we learn from these publications about how to improve study design to better understand the possible relationship between VR, gender, and cybersickness?

4.2 Phase 2: Scoping

After collecting data on approximately half of the papers, we revisited our categories to discuss preliminary findings. Phase 1 was an attempt to better understand possible factors influencing cybersickness during a study. However, in our review it became apparent that a clear connection cannot be drawn between, for example, headset type and increased cybersickness for women due to the number of confounds and the sheer variety of variables, and/or the lack of standardization and/or reporting across studies. Other categories, such as the number of dropouts, offered limited reports of data, or none at all. Given our intention to offer preliminary study design recommendations, we honed our set of categories again. Categories relating to technology type; exposure; study samples; consideration of gender; and measures of sickness were reported more reliably. We therefore focused on these latter categories for the remaining analysis in an effort to identify more generalizable criteria for studying the relationship between gender and cybersickness in VR. All investigated categories are listed and discussed in Section 5.2.

5 RESULTS

In this section we provide commentary on the current state of research pertaining to gender, cybersickness, and VR. In Section 5.1, we discuss the survey and review papers, which for the most part complicate understandings of gender and cybersickness by repeating inconclusive findings. In Section 5.2, we discuss the study-based papers, beginning by identifying a number of possible confounds, before moving to the categories that more directly report on possible connections between gender, cybersickness, and VR. We will contextualize these results in Section 6, and their implications in Section 7.

Survey papers that qualified per our criteria (12) are summarized in Section 5.1 and span from 2000 to 2020. Of these studies, five were published in the years 2015, 2017, 2018, 2019 (noting that our search ends at February 2020). After a discussion of the survey papers (12), in Section 5.2, we turn to a detailed analysis of the studies evaluated (59), which span the years 1996 to 2019. The oldest paper in our original set prior to screening was from 1994. We begin our analysis of the studies with the high-level results from some of the categories explored in Phase 1 before moving into those that came into focus in Phase 2 (Section 4.2).

5.1 Survey Papers

Of the publications deemed eligible for the final stage of our systematic review, 12 of the 71 were categorized as survey or review papers, with the remaining 59 classified as studies. In this section, we analyze the surveys and reviews to better understand synthesized perspectives on the possible relationship between cybersickness, VR, and gender. On average, these papers surveyed 246 manuscripts from databases such as PubMed, IEEE, ACM, and Google Scholar, in either a literature review format, theoretical discussion, or systematic review. Each of the 12 papers approached cybersickness differently, with topics including measuring motion sickness [44], possible correlations between migraine symptoms and cybersickness symptoms [67], technology-aided psychotherapy [21], the optimal visual modality in VR [86], and the representation of female authors and participants in VR research [68]. While gender and cybersickness are mentioned in each paper, the possible relationship between the two is not always considered, and the possible causes and mitigating factors are inconclusive or overlooked entirely.

The most common hypothesis presented as a possible reason for gendered cybersickness is discussed in 3 of the 12 papers [20, 44, 47], which refer to a reported difference in field-of-vision (FOV) of female participants. Women are said to have a wider FOV, which in turn is said to increase susceptibility to simulator sickness in VR due to a likelihood of flicker perception. A simpler, more obvious explanation is presented by Peck et al. [68], whose analysis of studies from the IEEE VR conferences from 2015-2019 showed both an under-representation of female participants and female authors. To ascertain the possible bias caused by this under-representation, the authors performed a subsequent meta-analysis of 21 papers to assess how female author or participant representation might affect results. The authors conclude that "smaller increases in simulator sickness following VR exposure were observed in studies with a greater proportion of female participants," suggesting that female participants' presumed susceptibility to cybersickness may simply be due to biased study design. While the authors note a number of limitations with this finding (e.g., challenges comparing variables across studies), they offer the reminder that "conclusions drawn from samples with inadequate gender diversity may not accurately characterize simulator sickness in the general population." The remaining survey and review papers [15, 21, 35, 52, 57, 67, 86, 93] mention gender as a possible factor of cybersickness, but do not discuss how or why a gendered difference might occur.

Overall, however, as with our exploratory review of CHI2019 VR papers, there are discrepancies and unknowns about how gender is understood. For example, for most of the reviews it is unclear whether any consideration is made for trans or non-binary identities. The one paper that mentions non-binary gender identities is the review by Peck et al. [68], which specifically excludes considerations for non-binary identities in their review of author gender, and excludes the 1 non-binary participant out of the 9,557 participants across 319 studies in order to make comparisons across male and female participants. The authors conclude with recommendations for more representative samples of participants, and more accurate reporting of demographics. Our own review confirms that such data gathering and reporting practices are needed across the vast majority of VR research.

5.2 Study Papers

In this section, we discuss our two phases of analysis for the 59 study papers revealed in our survey. In the first phase, we identify a number of possible confounds that are often unacknowledged when attempting to extrapolate findings across VR research. In the second phase, we hone the categories of analysis in an attempt to identify more consistent variables.

5.2.1 Phase 1: Categories with High Variability. While investigating and analyzing our set of 59 study papers, we began to note the possible confounding factors of cybersickness due to the many variables across system design and study design. The papers examined various aspects of VR; some noted the causes and symptoms of cybersickness as experienced by their participants [4, 40, 83] while others did not. The studies showed wide variation in terms of system design and study design, including variations in the amount of exposure time, break time, and other variables. The variety of technology used in research on VR creates a large amount of experimental data to disentangle, making it difficult to ascertain how and why cybersickness occurs and what needs to be done to mitigate its effects. Although we do not attempt to hypothesize the causes of cybersickness, it is important to highlight the possible confounds based on this work and other previous literature. By identifying possible confounds, future research can begin to manipulate variables in order to better understand the relationship between gender and cybersickness.

Technology types. The variations in terms of hardware, software, and content is not only underreported, but also difficult to assess. From the papers we analyzed, the most common displays used were head-mounted devices (HMD) (e.g., [32, 63, 69]). Other display types include CAVEs (e.g., [27, 37]) and screen monitors [19]). Some of these display types were used as part of driving simulation experiments (e.g., [27, 38]). Driving simulations were seen being used in both HMD environments [50] and CAVEs [27]. The types of programs used in the 59 studies varied from being either custom-created for the experiment (e.g., [2, 9]), or commercially available to the public (e.g., [1, 16]). In some cases, the authors chose the display type in order to prevent other confounds such as the weight of an HMD, which could have caused participants to perform fewer head movements [42]. While all of these types of virtual experiences may be categorized as forms of VR, they each offer significantly different experiences, meaning that generalized claims across these studies is limited.

Exposure and break duration. The duration of VR exposure in the papers we analyzed varied in the specific type and amount of time participants had to be in the virtual environments. Some studies divided their research into blocks of time where participants were exposed in short bouts such as 2-10 minute intervals [55], while others asked participants to complete tasks within the experiment and did not clearly mention the amount of time exposed [99]. Experiments ranged from intervals of a couple of minutes [69] to 20 minutes or more [42]. Experiments also ranged in the duration of breaks between exposure; some studies reported breaks that lasted a couple of minutes or 10 minutes [60], while others did not report the duration of their break times at all [25]. While differences in exposure time and breaks may be necessary depending on the study design, these examples show that there is too much variation to assess the direct effects on cybersickness results.

Independent and Dependent Variables. The studies that we analyzed also varied greatly in the dependent and independent variables they measured. Independent variables included factors such as latency [94], navigation [42], level of joystick control [72], furnishing of virtual environment rooms [9], etc. Dependent variables ranged from the rate of control [42], to simulator sickness [19], to spatial comprehension [9] and more. Some papers, such as Nichols [64], call on the need to recognize virtual reality induced symptoms and effects (or VRISE) as a multi-factorial problem. Sex was considered as an independent variable in two studies [1, 19], but gender and sex were more likely to be reported as demographic data. There is also a large variety in the way in which gender and sex are considered as part of the demographic data, with no clear consensus in the way gender was collected, reported, or analyzed.

Summary. It can be difficult to ascertain how this variety of factors may influence cybersickness, including displays, additional technologies, program type, exposure time, break time and variables studied. There are generally few references to these variables in existing studies, and we do not have sufficient control to identify clear relationships between a single factor and susceptibility to cybersickness. The papers we analyzed did not report these factors as part of their findings, making it difficult for us to make clear recommendations on how to mitigate cybersickness. For each of these categories (technology display, custom vs. commercial, exposure time, and break time), there is too much variety to be able to conclusively say what factors contribute to cybersickness. This lack of data is exacerbated by a lack of reporting on gender. Further research is needed to better understand how these possible confounds may contribute to cybersickness symptoms and onset.

5.2.2 Phase 2: Categories with Broad Patterns. In our second phase of analysis, our goal was to provide a more in-depth review across all 59 study papers by identifying categories that might offer more consistent reporting on the possible relationship between cybersickness and gender.

Gender and Sex. Perhaps the most obvious category to the topic at hand is how the studies in our systematic review wrote about, treated, and reported on participant gender, or in many cases, sex. Sample sizes ranged from 10 to 1102, with a mean of 122 participants and a median of 50. Reporting on basic participant information was not always straightforward: some papers required reverse engineering to determine the number of women included (e.g., from percentages of women provided across multiple study groups with a given N [91]). This could be attributed to differing disciplinary expectations, however, a lack of clear reporting can lead to confusion: one paper reports a sample of 60 participants with 39 (or 65%) males, but their limitations state that "After careful deliberation, the decision was made to recruit males only due to disproportionate attrition" [60] (in this case, we assume an error in transitioning from reporting on the breakdown of the original 122 participants recruited, to just those 60 remaining post-exclusion). Two papers [50, 51] did not report on the number of women in the sample, although one noted that groups were "balanced," which we assume to be 50%. The average number of women included in these studies was 54, with a range of 0 to 467, and a median of 24. This amounts to approximately 47.5% female representation across all of these studies, noting that inclusion ranged from 0% (all men in one study, [60]) to 100% (all women, in two studies [33, 76]).

Of the studies we analyzed, none made statements demonstrating any consideration for non-binary or transgender participants either in passing reference or in their approach to data collection or analysis. We also failed to detect any detailed information on *how* gender information was collected by researchers, either through statements by the authors, or implied through phrasing (e.g., "participants identified themselves as..."). We are therefore unable to ascertain whether gender or sex were ascribed by the researchers, or gathered through self-report, and what options were available for participants to self-identify.

Primary or Supplementary. When categorizing papers, we used "Primary" or "Supplementary" to denote whether the study foregrounded gender as a factor to be studied (i.e., it was included in their experimental design and/or constituted part of the overall research questions). In "Supplementary" cases, gender was analyzed in post-hoc analyses and/or was not considered a main factor of study. Out of the 59 studies included in this review, only 11 considered gender as a primary factor. The vast majority of papers (45, or 76%) included gender as a supplementary finding. As mentioned above, the three remaining studies had single-gender samples.

Difference found? Of those studies in which gender was a primary focus (excluding the studies with homogeneous samples), most found gender differences in their analysis (e.g., [2, 32, 69, 72, 93]). However, the proportion of studies finding a difference varies depending on whether "difference" is defined as something for which the authors were looking for with regards to their main research question(s), or if they found differences in other measures: most, if not all, reported gender differences in areas like dropouts and reports of sickness. Four studies reported no differences, but again, included commentary on dropouts and sickness, making it difficult to determine what does and does not constitute "difference found." One study reporting no gender differences also reports having just one female participant remaining after 10 minutes [63], so making any claims about difference would not be possible given the sample. We also found that the vast majority of studies where gender was a supplementary consideration reported gendered differences in experience, pointing to the value of including gender as a consideration in VR studies regardless of whether the study is 'about' gender.

Measures of Cybersickness. From our data set, 48 out of 59 papers (81%) used the Simulator Sickness Questionnaire (SSQ) [36] to measure participants' symptoms. Of these 48 studies, 9 used additional supplementary measures to capture sickness data: general questionnaires made by the authors (3), MSSQ/MSSQ-short (4), Body Awareness Questionnaire (1), VIMSSQ (1), FMS (3), malaise rating (1), MHQ (2), SS-VAS (1). This means that 39 used the SSQ alone, and 9 used some combination of SSQ plus other measures. Other studies used standalone measures to capture cybersickness information: one study used the MSSQ-S alone [51]; one used MHQ alone [75]; or verbal feedback alone [55]. Overall, while every one of the 59 studies measured sickness in some way, the differences across these various measures of cybersickness makes it difficult to develop generalizations about how, when, or why cybersickness occurs across these studies. Comparing and compiling differences across these measures is another opportunity for future work.

Dropouts. Dropouts were a concern across all studies surveyed: there appears to be a general understanding that where there is virtual reality, there will be cybersickness, and where there is cybersickness, there will be individuals unable to continue participating. How this reality is handled varies. Some studies (e.g., [42]) reported on the number of participants, their gender, and reasons for discontinuing participation in the study; others reported on only one of or some combination of these values. Cybersickness was the most frequent cause for terminating an experiment, although some other reasons included equipment malfunction, and participants not following instructions. Women were consistently recorded as more likely to discontinue participation in the experiment due to intensity of cybersickness [19, 29].

Some papers retained the data for dropouts and reported on it to the extent that was possible [37, 43, 72], providing important context on the full spectrum of participant experiences. One study had one female participant drop out and noted "shortly after exiting the VE [virtual environment], she induced vomiting and vomited three times. This participant reported that she has a long history of motion sickness in both cars and airplanes. During the follow-up call, she reported that she started feeling better about two to three days after the session and stated that her experience in this study was probably [her] worst case of motion sickness ever" [43, p. 1513]. Others (e.g., [3]) sought new replacement participants with similar demographics and scoring on any pre-study measures: the implications of this practice are discussed in Section 6.

Demographics, inclusion, exclusion. The demographics of the participants that were reported varied greatly across studies. All studies except for two included gender as part of their participant sample breakdown [50, 51]. Age was commonly noted, and the average age pool was composed of college students (e.g., [82, 99]). Papers rarely noted the race or ethnicity of the participants being studied. When race or ethnicity was noted, it was still not tested or discussed [60]. The level of prior video-gaming or VR experience of the participants was also commonly noted among studies (e.g., [7, 31, 42]). Although these types of factors were sometimes included, there were infrequent tests to explore whether differences exist between the various demographic data of their participants.

Of the papers we reviewed, only some noted the exclusion criteria or why they chose the criteria that they did. Some papers noted that participants who dropped out were excluded from final analysis but do not state why participants dropped out, while others mention that participants dropped out because of the cybersickness they experienced (e.g., [53, 83]). The inclusion criteria of the papers investigated were factors around how participants might not be able to take part in the study, such as their vision. For example, some studies asked participants to complete questionnaires to investigate if they had normal or corrected vision, or asked participants to complete a vision test [32, 94]. Papers were overall more likely to report on exclusion criteria rather than inclusion criteria, however many papers did not explicitly report these criteria at all.

6 DISCUSSION

In our analysis, we focus on the study papers evaluated and discuss our findings in the context of their implications for future research on gender and cybersickness in VR. In particular, we spend time on those variables that presented more identifiable patterns and opportunities for researchers doing studies in VR. When appropriate, we contextualize these factors within broader patterns in HCI research. Although we critique particular examples in the following, it is not our intention to target individual researchers, but rather to provide examples that demonstrate what we view as a systemic issue across the literature we reviewed.

6.1 *How* is gender involved in these studies?

6.1.1 How is gender treated by the authors? One thing the CHI community already has to its advantage based on our exploratory review is that it reports fairly reliably on gender breakdown of participants, although somewhat imperfectly: 65 of the 96 papers surveyed in that review (Section 3) reported on participant gender. The manner in which participant breakdowns were reported in our review of literature showed inconsistencies as to where, how, and whether this information was even provided. Since the papers in our review represent a variety of domains, we are unable to comment on whether their use of the terms "gender" or "sex" is the result of different disciplinary standards or attributable to other factors, such as the authors' understanding of the gender binary being projected through their reporting. There are existing publications within CHI that discuss implications of, and provide recommendations for, this type of reporting (e.g., [10, 11, 74]). More intentionality is needed in the use of these terms in order to avoid essentializing claims that can add uncertainty around how this data was collected, and what was inferred by the authors as a result. Given that the HCI literature we reviewed largely ignored the relationship between gender and cybersickness, we highlight the opportunity to adopt more inclusive understandings of gender in VR studies. For example, Burtscher and Spiel [11] broadly identify three common social understandings of gender as *essentialism*, *performance*, or *identity*. Although all three conceptions were present in our review, the most common was the essentialist perspective.

In one example of the *essentialist* view, the authors align menstruation with womanhood [16]. Despite their focus on this biological process, the authors still use "gender," demonstrating a biologically deterministic view whereby the association of menstruation with a particular sex implicates gender [11]. While research about potential impacts of hormones on susceptibility to cybersickness can add value by examining additional variables, it implicitly positions men as the control group with "normal" experiences in VR, and the essentialized category of women as the deviant group to be ameliorated. Rather than adjusting systems to people, people are forced to adjust to systems. After completing our systematic review, we noticed that this approach is still being applied in 2021. For example, Lim et al. [48] use an all-male participant pool to eliminate hormonal fluctuations associated with menstruation. They selected participants who were "healthy and easy" to participate in the experiment with the stated goal of clarifying and quantifying factors causing cybersickness.

An example of the *performative* view of gender can be seen in a paper that claims that visually induced motion sickness (VIMS) is more severe and has a faster onset for women (coupled with a 49% dropout rate, compared to 18% among men): despite having tested selected variables in a three-factorial between-participants design, the authors disregarded this outcome, writing "the reason why VIMS is more prevalent in women is not yet understood ... One reason, however, may be a cultural difference rather than a true sex-effect or physiological difference, with females possibly being more open-minded about their feelings than males" [19]. In this case, such a claim undermines what the authors studied, analyzed, and published by attributing their results to stereotypes.

Keyes [39] highlights how externally defining aspects of participants' self-representation can amplify pre-existing inequalities, and in turn, thwart a person's self-determination by denying their *identity*. There are multiple sources published in and intended for the HCI community to support individual self-determination through thoughtful and informed engagement with gender throughout the research process (e.g., [74, 80]), and these understandings of gender draw on decades of research across feminist, queer, and trans studies. However, none of the studies included in our systematic review made reference to gender beyond the binary, nor did they demonstrate recognition of the changing nature of gender itself within a broader social or cultural context. If the option to self-identify was included in study protocols but not reported, then it is not possible to judge whether erasure has occurred. Reporting language can help to clarify that participants have had a chance to self-identify, e.g., "7 [participants] identified themselves as females, and 5 as males" [6]. In VR research, there is a clear gap in reporting what options were available to participants, whether or not they were selected, and indicating through word choice who established the gender of participants.

6.1.2 *Primary vs. Supplementary Framing of Differences Found.*

When screening the papers for this review, it was not immediately apparent which papers would consider gender a primary focus. We were then surprised at the low proportion of "Primary" papers in our set compared to "Supplementary": while gender effects were framed as incidental findings in the latter group, each of these papers demonstrates a need to consider gender at the forefront of the study design. Our search criteria biases our sample of papers towards those that have included gender or sex as a relevant term in their publication, meaning that we have not evaluated papers that make no mention of it at all. Whether or not gender differences in cybersickness were part of the authors' own research questions, we found that the mere act of reporting on gender provides valuable information for other researchers. 45 of 59 (76%) studies in our review considered gender as a supplementary reported result. Of these 45 studies, 26 of them (57%) reported these differences as relating to cybersickness. Had a consideration and reporting of gender been the norm rather than the exception, perhaps a meta-analysis would have been possible for this systematic review. Going forward, we encourage authors to take away an increased awareness of the potential relationship between gender and cybersickness, and to report their findings even when their VR studies are not 'about' gender, so that future meta-analyses may be conducted.

Over the course of our review, we were able to find authors who expanded on the potential impact of other individual factors as contributors to cybersickness. We note this finding to highlight the fact that gender is not likely to explain away all of cybersickness, and that future work should remain sensitive to the complexity of individual differences contributing to experiences in VR as opposed to essentializing based on categories of people. Wilson and Kinsela [94] contend that while women are reported to be more susceptible to sickness, differences may be due to individual susceptibility. For example, this is demonstrated in Nguyen et al. [63]'s analysis of the ability to detect curvature redirection, wherein men on average performed better on detecting curvature gain, but there was high variance *within* their gender groups, leading the authors to question the importance of visual dependence among individuals rather than groups. Chen et al. [13] also find large individual variation in susceptibility to sickness, while also noting several of the same confounding factors that were considered in this review. While controlling for the effect of susceptibility and citing Barnett-Cowan et al. [5], the authors determine that gender *was* indeed a contributing factor not to be overlooked. Overall, a high degree of variation within two binary gender categories throws into question whether binary gender is a *good* choice or a *convenient* choice for data collection and analysis, particularly in light of other individual factors emerging such as visual dependency or susceptibility. Given the degree of uncertainty in the data and disagreements across publications, we contend that it is necessary to continue investigating the effects of gender alongside other identity factors on cybersickness. Reporting on this data is necessary within diverse contexts and research settings, and across a range of possible variables.

6.2 *What* is being studied?

6.2.1 *Measures: Can the diversity of cybersickness be adequately, or satisfactorily, quantified?*

The most popular measure of cybersickness among all papers was the Simulator Sickness Questionnaire (SSQ). As mentioned in Section 5, every study in our sample measured cybersickness in one way or another. In contrast, only 21% of the CHI 2019 papers reported measuring sickness in some way. Regardless of proportions, what we have learned about the vastness of independent differences would suggest that a plurality of approaches would be required to capture these various nuances. A feminist approach to data science reminds us that "what gets counted counts" [18]; accordingly, we raise concerns about the repeated exclusion of trans/non-binary identities, and the continuing reliance on and validity of the SSQ as such a singularly dominant approach to document cybersickness.

From our sample of papers, it became obvious that susceptibility and sickness co-occurred in the reporting. Authors intended to contribute variables that served as predictors of susceptibility to cybersickness (e.g., [79]); there were studies which stated that women have higher susceptibility than men in the framing of their work [30, 43, 59], implying a trend that "susceptibility" and "women" are tightly coupled concepts. It is therefore concerning that some papers used susceptibility to motion sickness as exclusion criteria for their studies [60], although others required that participants demonstrate some level of susceptibility [16]. The role or relevance of susceptibility altogether was brought into question by Weech et al. [93] who found differences in cybersickness but not in reports of susceptibility documented using the MSSQ.

There is an opportunity to better identify the *contexts* in which these measures are developed and applied, and what is actually explainable from the outputs. One paper explained that an SSQ score above 20 meant that participants were "sick" [27], whereas the intended interpretation is that the simulator itself is "bad" [36]. "Sick" is a feeling, not a number, despite the necessary efforts to quantify that feeling. Other measures used such as the Motion History Questionnaire (MHQ) and Motion Sickness Susceptibility Questionnaire (MSSQ) both take past experience as predictors of whether participants will be impacted by cybersickness. Rather than filtering for susceptibility, a major stepping stone in cybersickness research would be to move from aiming to forewarn those who are susceptible (e.g., [85]) to determining what specific design elements (e.g., hardware, software, interaction techniques) are needed for a system that is enjoyable for everyone to use.

Recent research in VR has begun to develop and adopt more targeted approaches to assessing and evaluating cybersickness, including the CSQ [87] and the VRSQ [41]. Our review suggests that not only is there an ongoing need to iterate on the tools we use to better understand and mitigate cybersickness, but also a need to examine how these tools are being used, and how gender is considered before, during, and after data collection. The current lack of data on gender that our study demonstrates raises important questions for future research. This includes a need for more generalizable best practices for VR research, as well as a need for practical case studies developed with rigorous attention to how the data is collected, and who is represented by that data.

6.2.2 "No one verbally complained". We noticed a pattern of little qualitative feedback being reported to contextualize participants' experiences. Some authors expressed a level of confusion when SSQ scores indicated high levels of sickness, but "no subject complained" [4] (we find a similar statement in [92]). Asjad et al. [4] state that it was difficult to assess the severity of symptoms, since "no one verbally complained." Short statements like this indicate that authors across our systematic review were not in the practice of including semi-structured, debriefing, or exit interviews. It also indicates that had they taken this step to solicit qualitative feedback, they may have been able to more accurately contextualize their findings using the verbal feedback collected.

In our review, we encountered essentializing claims regarding gender and the expression of cybersickness or discomfort in VR. We emphasize that the value in providing qualitative feedback from participants is not to add further anecdotal evidence to the "women are/are not more likely to express discomfort" debate, but rather to provide the information beyond what a questionnaire is capable of capturing. If the questionnaire (whether the SSQ or another measure) is adequately capturing the breadth of experiences, then this would not be necessary; however, given the number of open questions following this review, it is clear that this is not the case. Questionnaires like the SSQ tell us that people felt sick, but do not tell us how to *mitigate* that feeling. Supplementing quantitative measures with qualitative methods (not only interviews, but observational and other methods) can provide more detail on how, why, and when sickness is induced.

7 RECOMMENDATIONS & LIMITATIONS

Our systematic review revealed that there is a large amount of information and detail missing about how gender relates to cybersickness, and how to mitigate the cybersickness experienced by women. Our review calls attention to these unknowns, and the need to manipulate these variables to study them more in-depth. In this section, we provide some preliminary recommendations for future VR research. It should be noted that there is still a need for more generalizable best practices within VR research in HCI that take into account the realities of various sample sizes, budgets, etc. As more research and information becomes available on the relationship between gender and cybersickness, these recommendations will need to be modified and updated. However, these recommendations provide a much-needed step towards maintaining consistency, replicability, and effective VR research progression benefiting all genders.

Our preliminary recommendations include:

(1) **Clear reporting of demographic characteristics:** Although we focus on the lack of reporting around gender, it is clear that there is also a general lack of reporting around other demographic characteristics, such as race, ethnicity, and age. More detailed reporting and more diverse participant pools are necessary to ensure more representative and generalizable findings [68]. In order to ensure that VR is safe for everyone, future work will first need to assess whether reports of gendered cybersickness are due to biological factors, social factors, some combination of the two, and/or misrepresentations due to a lack of data. Simply put, the underrepresentation of women as participants in VR studies [68] calls existing theories into question.

(2) **Broader considerations of gender:** Beyond representation, our survey reveals that there is an implicit lack of reporting around what is assumed (i.e., biologically, socially) about gender. It is incumbent on researchers and reviewers to develop considerations of gender beyond a binary, and to develop up to date practices of reporting on gender that does not essentialize and assume. Decades of feminist, queer, trans and other scholarship offers many examples of best practices, and more recently research in HCI offers domain-specific considerations [74, 80]. Importantly, these understandings of gender must come from within communities that are most impacted by the results, with greater care to work with and/or position them as experts of their own bodies and experiences. A lack of representation from authors to participants [68] can lead to biased findings.

(3) **Clear reporting of drop-outs:** This includes the reporting of drop-out demographics (such as gender), at what point they ceased participation in the study, and the reasons as to why each participant dropped out (e.g., experience of cybersickness). Having adequate and detailed reports of

drop-outs will enable future research to disentangle the relationship between gender and cybersickness, and elucidate under-explored factors related to more extreme responses. As ethical guidelines stipulate that participation is voluntary, in some cases it may be impossible to fully understand why some participants drop out of a VR study. It may also be the case that ethics protocols will specify that drop-out data must be discarded; due to the potential value of this data, there may be an opportunity to reconsider what aspects of the data can be included in cases where participants drop out due to cybersickness. This lack of data must be acknowledged, and should also be a reminder to develop practices to safely account for all experiences.

(4) **Study design:** Participant exclusion criteria was not frequently stated in the papers we reviewed, which results in a lack of understanding of why some participants are not represented in research. Documenting exclusion criteria is especially important to better understand who and what is considered from the outset. Due to the possibility of cybersickness, any study that uses VR must find ways to carefully choose a spectrum of measures that are sensitive to differences in experience. A one-size-fits-all model in system design or study design may not be adequate or inclusive. For example, because of the variety of factors that may cause cybersickness, researchers may need to carefully consider the duration of study conditions as it relates to average time of onset for symptoms [58, 73], time buffers between conditions as symptom onset can occur within 24 hours of exposure [85], as well as how and when they use between- or within-participants study design to account for adaptation effects [22]. Future research could provide more detail regarding the content and experimental setup, i.e., diagrams, source code, product information.

(5) **Measuring cybersickness:** There was a wide variety in the type of measurements that researchers used to assess cybersickness. Differences across measurements can make it challenging to formulate concrete themes across diverse research areas, but these questionnaires show that cybersickness remains a clear and ongoing problem in VR research. At the very least, future research can make use of a validated cybersickness questionnaire like the SSQ, but also supplement it with qualitative data such as verbal feedback and report differences (and absence of differences) based on gender and other demographics. This is also an opportunity for researchers to critically reflect on the possible limitations of these methods and/or to ascertain whether newer measurements or validated questionnaires may be more appropriate. Nevertheless, documenting cybersickness among participants provides information that, while not always directly relevant to the primary research questions, relates to the broader issue affecting all research in VR. More work is needed to better understand both the onset and post-exposure effects of cybersickness.

A closer reading of the work presented in our review demonstrates that complexity introduced by the study design may limit researchers' ability to derive meaning from their data with regard to cybersickness, which in turn may make it more difficult to extrapolate findings for future work. An overarching recommendation for future research is to adopt an interdisciplinary view that strives to more broadly address some of the key concerns raised by our analysis. As VR impacts so many fields it may be increasingly necessary to search across disciplines to better acknowledge the work that has been done, and the work that still needs to be done.

7.1 Limitations

The limitations of our review are broad. Some of these limitations are due to the methodology of conducting the systematic review, while others pertain to the perspective we took while conducting the review itself. In this section, we discuss how the need to contextualize the work reviewed during this project came with its own constraints due to the irregularity in the data collected, and we address the challenges in filling in the missing gaps of the relationship between gender, cybersickness, and VR.

7.1.1 *Limitations of our method/approach.*

Search bias. In our methodology, we intended to select papers that had gender as part of their work. Our approach therefore creates its own bias within the sample of papers that we based our review on. For example, a number of papers had near-gender parity, but this is likely due to our choice of gender as inclusion criteria rather than what is representative of VR research more broadly. The same can be said of our inclusion of cybersickness and simulator sickness. As both cybersickness and simulator sickness were part of our search strings, a high number of the resulting papers therefore measured these concepts in some way. Compared to the CHI 2019 papers from our exploratory review, it is unsurprising that our systematic review yielded many more papers that were about cybersickness and/or simulator sickness.

Databases. A meta finding related to our technique is in the instability of the databases themselves. Using the university library, with the same search string on the same day, would yield different results (in terms of the number of records). Library staff could only offer that the results not showing up were duplicates of the same records from various sources. Similarly, we had to adjust our search methods to suit different databases, and small changes would yield different results. This brought into the foreground the things that remain out of researchers' control when trying to collect data through systematic reviews, which we felt was worth reflecting on for other authors who may perform similar techniques in the future.

7.1.2 Limitations of our perspective. Our findings emphasize a lack of data with regard to the relationship between gender and cybersickness, and the lack of consensus and possible confounds in the way VR research is reported. However, in our attempt to highlight these unknowns, we inadvertently reinforce the concept of binary genders through our focus on the comparisons "between" genders. The direct causes behind the potential difference between the susceptibility of cybersickness among different genders might not even be due to underlying biological differences. Research has noted that the true causes are still unknown, and although biological aspects such as hormonal cycles are a possible factor, as we discuss, there is still not enough evidence to validate these findings. There are also an abundance of other factors beyond gender that we are unable to address that face similar problems. For example, elderly participants are often excluded because they are also said to have an increased susceptibility to cybersickness compared to younger participants. However, there are still some studies that choose to include senior participants, and the relationship between age and susceptibility did not appear to have a clear consensus in our review. Similarly, race and ethnicity are often overlooked, and like gender there is anecdotal evidence that such demographic characteristics may also have an effect (e.g., [33]), though we stress that it is important not to extrapolate results based on limited findings. Lastly, participants with disabilities are often excluded from VR research. For example, locomotion studies in VR that involve a treadmill or walking around a room, might assume or require an ease of physical mobility. Additionally, low vision is a common reason for exclusion from VR studies [79]. Each of these factors underscores the need for more diversity within participant samples, and to increase the demographic pool of participants as a whole within VR research.

8 CONCLUSION

Our systematic review is motivated by studies that report a gendered susceptibility to cybersickness across VR research. After conducting an exploratory review within CHI 2019 proceedings and papers, we conducted a broad, interdisciplinary review of research that comments on gender, cybersickness, and VR. Our work reviews survey papers as well as study papers, and we identified a number of confounding factors as well as under-reporting across categories with a direct impact on gender and cybersickness. We identified several gaps in the research pertaining to gender, as well as other demographics and factors such as dropout, exclusion criteria, and study design. We found that papers in our review primarily treat gender and cybersickness as a secondary aspect of their research and do not analyze the data or provide meaningful recommendations for mitigating the effects of cybersickness. Our analysis contributes insight into a decades-old problem in VR research, as well as preliminary recommendations for how to conduct future VR research. Our overarching recommendation is that considerations for gender and cybersickness are important at every stage of VR research, from study design, to data collection, to analysis.

ACKNOWLEDGMENTS

Thank you to Licheng Zhang and Marvin Pafla, the NSERC Postgraduate Scholarship program, the University of Waterloo's Touchlab, EngHCI group, and the Games Institute. This work was made possible by NSERC Discovery Grant 2016-04422, NSERC Discovery Accelerator Grant 492970-2016, NSERC CREATE Saskatchewan-Waterloo Games User Research (SWaGUR) Grant 479724-2016, and Ontario Early Researcher Award ER15-11-184.

REFERENCES

[1] Majed Al Zayer, Isayas B. Adhanom, Paul MacNeilage, and Eelke Folmer. 2019. The Effect of Field-of-View Restriction on Sex Bias in VR Sickness and Spatial Navigation Performance. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*. ACM Press, Glasgow, Scotland Uk, 1–12. Retrieved 09/15/2020 from http://dl.acm.org/citation.cfm?doid=3290605.3300584.

[2] Cassandra N. Aldaba and Zahra Moussavi. 2020. Effects of virtual reality technology locomotive multi-sensory motion stimuli on a user simulator sickness and controller intuitiveness during a navigation task. *Medical & Biological Engineering & Computing*, 58, 1, (January 2020), 143–154. Retrieved 09/15/2020 from http://link.springer.com/10.1007/s11517-019-02070-2.

[3] Laura Arns and Carolina Cruz-Neira. 2004. Effects of physical and virtual rotations and display device on users of an architectural walkthrough. In *Proceedings of the 2004 ACM SIGGRAPH international conference on Virtual Reality continuum and its applications in industry - VRCAI '04*. ACM Press, Singapore, 104. Retrieved 08/27/2020 from http://portal.acm.org/citation.cfm?doid=1044588.1044608.

[4] Noorin Suhaila Asjad, Haley Adams, Richard Paris, and Bobby Bodenheimer. 2018. Perception of height in virtual reality: a study of climbing stairs. In *Proceedings of the 15th ACM Symposium on Applied Perception*. ACM, Vancouver British Columbia Canada, (August 2018), 1–8. Retrieved 09/14/2020 from https://dl.acm.org/doi/10.1145/3225153.3225171.

[5] M. Barnett-Cowan, R. T. Dyde, C. Thompson, and L. R. Harris. 2010. Multisensory determinants of orientation perception: task-specific sex differences: Sex differences in orientation perception. *European Journal of Neuroscience*, 31, 10, (May 2010), 1899–1907. Retrieved 09/30/2016 from http://doi.wiley.com/10.1111/j.1460-9568.2010.07199.x.

[6] Joanna Bergström, Aske Mottelson, Andreea Muresan, and Kasper Hornbæk. 2019. Tool extension in human-computer interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (CHI '19). Association for Computing Machinery, Glasgow, Scotland Uk, 1–11. https://doi.org/10.1145/3290605.3300798.

[7] Maximino Bessa, Miguel Melo, David Narciso, Luís Barbosa, and José Vasconcelos-Raposo. 2016. Does 3D 360 video enhance user's VR experience?: An Evaluation Study. In *Proceedings of the XVII International Conference on Human Computer Interaction - Interacción '16*. ACM Press, Salamanca, Spain, 1–4. Retrieved 09/12/2020 from http://dl.acm.org/citation.cfm?doid=2998626.2998669.

[8] Lindsay Blackwell, Nicole Ellison, Natasha Elliott-Deflo, and Raz Schwartz. 2019. Harassment in social virtual reality: challenges for platform governance. *Proc. ACM Hum.-Comput. Interact.*, 3, CSCW, Article 100, (November 2019), 25 pages. https://doi.org/10.1145/3359202.

[9] Sabah Boustila, Antonio Capobianco, and Dominique Bechmann. 2015. Evaluation of factors affecting distance perception in architectural project review in immersive virtual environments. In *Proceedings of the 21st ACM Symposium on Virtual Reality Software and Technology - VRST '15*. ACM Press, Beijing, China, 207–216. Retrieved 09/14/2020 from http://dl.acm.org/citation.cfm?doid=2821592.2821595.

[10] Adam Bradley, Cayley MacArthur, Mark Hancock, and Sheelagh Carpendale. 2015. Gendered or neutral?: considering the language of HCI. In *Proceedings of the 41st Graphics Interface Conference*. Canadian Information Processing Society, 163–170. Retrieved 10/18/2015 from http://dl.acm.org/citation.cfm?id=2788919.

[11] Sabrina Burtscher and Katta Spiel. 2020. "but where would i even start?": developing (gender) sensitivity in hci research and practice. In *Proceedings of the Conference on Mensch Und Computer* (MuC '20). Association for Computing Machinery, Magdeburg, Germany, 431–441. https://doi-org.proxy.lib.uwaterloo.ca/10.1145/3404983.3405510.

[12] Michelle L. Cadieux, Michael Barnett-Cowan, and David I. Shore. 2010. Crossing the hands is more confusing for females than males. *Experimental Brain Research*, 204, 3, (July 2010), 431–446. Retrieved 08/11/2016 from http://link.springer.com/10.1007/s00221-010-2268-5.

[13] Wei Chen, Jian-Gang Chao, Xue-Wen Chen, Jin-Kun Wang, and Cheng Tan. 2015. Quantitative orientation preference and susceptibility to space motion sickness simulated in a virtual reality environment. *Brain Research Bulletin*, 113, (April 2015), 17–26. Retrieved 09/17/2020 from https://linkinghub.elsevier.com/retrieve/pii/S0361923015000246.

[14] Bob Cheung and Kevin Hofer. 2002. Lack of gender difference in motion sickness induced by vestibular Coriolis cross-coupling. *Journal of Vestibular Research: Equilibrium & Orientation*, 12, 4, 191–200.

[15] S. Classen, M. Bewernitz, and O. Shechtman. 2011. Driving Simulator Sickness: An Evidence-Based Review of the Literature. *American Journal of Occupational Therapy*, 65, 2, (March 2011), 179–188. Retrieved 09/12/2020 from http://ajot.aota.org/Article.aspx?doi=10.5014/ajot.2011.000802.

[16] Stacy A. Clemes and Peter A. Howarth. 2005. The Menstrual Cycle and Susceptibility to Virtual Simulation Sickness. *Journal of Biological Rhythms*, 20, 1, (February 2005), 71–82. Retrieved 09/15/2020 from http://journals.sagepub.com/doi/10.1177/0748730404272567.

[17] Mary Czerwinski, Desney S. Tan, and George G. Robertson. 2002. Women take a wider view. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 195–202. Retrieved 09/21/2014 from http://dl.acm.org/citation.cfm?id=503412.

[18] C. D'Ignazio and L.F. Klein. 2020. *Data Feminism. Strong Ideas*. MIT Press. https://books.google.ca/books?id=Hu_KyQEACAAJ.

[19] Sarah D'Amour, Jelte E. Bos, and Behrang Keshavarz. 2017. The efficacy of airflow and seat vibration on reducing visually induced motion sickness. *Experimental Brain Research*, 235, 9, (September 2017), 2811–2820. Retrieved 09/15/2020 from http://link.springer.com/10.1007/s00221-017-5009-1.

[20] Simon Davis, Keith Nesbitt, and Eugene Nalivaiko. 2014. A Systematic Review of Cybersickness. In *Proceedings of the 2014 Conference on Interactive Entertainment - IE2014*. ACM Press, Newcastle, NSW, Australia, 1–9. Retrieved 03/11/2020 from http://dl.acm.org/citation.cfm?doid=2677758.2677780.

[21] Patricia R. DeLucia, Stephanie A. Harold, and Yi-Yuan Tang. 2013. Innovation in Technology-Aided Psychotherapy Through Human Factors/Ergonomics: Toward a Collaborative Approach. *Journal of Contemporary Psychotherapy*, 43, 4, (December 2013), 253–260. Retrieved 09/12/2020 from http://link.springer.com/10.1007/s10879-013-9238-8.

[22] Natalia Duzmanska, Pawel Strojny, and Agnieszka Strojny. 2018. Can simulator sickness be avoided? a review on temporal aspects of simulator sickness. *Frontiers in Psychology*, 9, 2132. https://www.frontiersin.org/article/10.3389/fpsyg.2018.02132.

[23] Chris Esposito. 1996. User interface issues for virtual reality systems. In *Conference companion on Human factors in computing systems common ground - CHI '96*. ACM Press, Vancouver, British Columbia, Canada, 340–341. Retrieved 09/14/2020 from http://portal.acm.org/citation.cfm?doid=257089.257358.

[24] Moira B. Flanagan, James G. May, and Thomas G. Dobie. 2005. Sex differences in tolerance to visually-induced motion sickness. *Aviation, Space, and Environmental Medicine*, 76, 7, (July 2005), 642–646.

[25] Cyrus K. Foroughi, W. Chris Wren, Daniela Barragán, Patrick R. Mead, and Deborah A. Boehm-Davis. 2015. Assessing Mental Rotation Ability in a Virtual Environment with an Oculus Rift. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 59, 1, (September 2015), 1849–1852. Retrieved 09/15/2020 from http://journals.sagepub.com/doi/10.1177/1541931215591399.

[26] Francesca C. Fortenbaugh, Sidhartha Chaudhury, John C. Hicks, Lei Hao, and Kathleen A. Turano. 2007. Gender differences in cue preference during path integration in virtual environments. *ACM Transactions on Applied Perception*, 4, 1, (January 2007), 6–es. Retrieved 10/27/2016 from http://portal.acm.org/citation.cfm?doid=1227134.1227140.

[27] Germán Gálvez-García, Marion Hay, and Catherine Gabaude. 2015. Alleviating Simulator Sickness with Galvanic Cutaneous Stimulation. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 57, 4, (June 2015), 649–657. Retrieved 09/14/2020 from http://journals.sagepub.com/doi/10.1177/0018720814554948.

[28] Omri Gillath, Cade McCall, Phillip R. Shaver, and Jim Blascovich. 2008. What Can Virtual Reality Teach Us About Prosocial Tendencies in Real and Virtual Environments? *Media Psychology*, 11, 2, (June 2008), 259–282. Retrieved 09/17/2020 from http://www.tandfonline.com/doi/abs/10.1080/15213260801906489.

[29] David A. Graeber. 2001. Use of Incremental Adaptation and Habituation Regimens for Mitigating Optokinetic Side Effects. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 45, 27, (October 2001), 1901–1905. Retrieved 09/14/2020 from http://journals.sagepub.com/doi/10.1177/154193120104502710.

[30] David A. Graeber and Kay M. Stanney. 2002. Gender Differences in Visually Induced Motion Sickness. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 46, 26, (September 2002), 2109–2113. Retrieved 09/17/2020 from http://journals.sagepub.com/doi/10.1177/154193120204602602.

[31] Aliya Iskenderova, Florian Weidner, and Wolfgang Broll. 2017. Drunk Virtual Reality Gaming: Exploring the Influence of Alcohol on Cybersickness. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM, Amsterdam The Netherlands, (October 2017), 561–572. Retrieved 09/16/2020 from https://dl.acm.org/doi/10.1145/3116595.3116618.

[32] Beverly K. Jaeger and Ronald R. Mourant. 2001. Comparison of Simulator Sickness Using Static and Dynamic Walking Simulators. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 45, 27, (October 2001), 1896–1900. Retrieved 08/14/2020 from http://journals.sagepub.com/doi/10.1177/154193120104502709.

[33] Jennifer T. T. Ji, Richard H. Y. So, and Raymond T. F. Cheung. 2009. Isolating the Effects of Vection and Optokinetic Nystagmus on Optokinetic Rotation-Induced Motion Sickness. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 51, 5, (October 2009), 739–751. Retrieved 09/14/2020 from http://journals.sagepub.com/doi/10.1177/0018720809349708.

[34] M. D. Jokerst, M. Gatto, R. Fazio, P. J. Gianaros, R. M. Stern, and K. L. Koch. 1999. Effects of gender of subjects and experimenter on susceptibility to motion sickness. *Aviation, Space, and Environmental Medicine*, 70, 10, (October 1999), 962–965.

[35] Marshall B. Jones, Robert S. Kennedy, and Kay M. Stanney. 2004. Toward Systematic Control of Cybersickness. *Presence: Teleoperators and Virtual Environments*, 13, 5, (October 2004), 589–600. Retrieved 09/12/2020 from https://www.mitpressjournals.org/doi/abs/10.1162/1054746042545247.

[36] Robert S. Kennedy, Norman E. Lane, Kevin S. Berbaum, and Michael G. Lilienthal. 1993. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology*, 3, 3, (July 1993), 203–220. Retrieved 03/04/2018 from http://www.tandfonline.com/doi/abs/10.1207/s15327108ijap0303_3.

[37] Behrang Keshavarz and Heiko Hecht. 2011. Validating an Efficient Method to Quantify Motion Sickness. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 53, 4, (August 2011), 415–426. Retrieved 09/15/2020 from http://journals.sagepub.com/doi/10.1177/0018720811403736.

[38] Behrang Keshavarz, Raheleh Saryazdi, Jennifer L. Campos, and John F. Golding. 2019. Introducing the VIMSSQ: Measuring susceptibility to visually induced motion sickness. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 63, 1, (November 2019), 2267–2271. Retrieved 08/28/2020 from http://journals.sagepub.com/doi/10.1177/1071181319631216.

[39] Os Keyes. 2018. The Misgendering Machines: Trans/HCI Implications of Automatic Gender Recognition. *Proceedings of the ACM on Human-Computer Interaction*, 2, CSCW, (November 2018), 1–22. Retrieved 01/15/2019 from http://dl.acm.org/citation.cfm?doid=3290265.3274357.

[40] Aram Kim, Nora Darakjian, and James M. Finley. 2017. Walking in fully immersive virtual environments: an evaluation of potential adverse effects in older adults and individuals with Parkinson's disease. *Journal of NeuroEngineering and Rehabilitation*, 14.

[41] Hyun K. Kim, Jaehyun Park, Yeongcheol Choi, and Mungyeong Choe. 2018. Virtual reality sickness questionnaire (VRSQ): Motion sickness measurement index in a virtual reality environment. *Applied Ergonomics*, 69, (May 2018), 66–73. Retrieved 09/14/2020 from https://linkinghub.elsevier.com/retrieve/pii/S000368701730282X.

[42] Ji-Sun Kim, Denis Gračanin, Taeyoung Yang, and Francis Quek. 2015. Action-Transferred Navigation Technique Design Approach Supporting Human Spatial Learning. *ACM Transactions on Computer-Human Interaction*, 22, 6, (December 2015), 1–42. Retrieved 09/15/2020 from https://dl.acm.org/doi/10.1145/2811258.

[43] Eugenia M. Kolasinski and Richard D. Gilson. 1998. Simulator Sickness and Related Findings in a Virtual Environment. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 42, 21, (October 1998), 1511–1515. Retrieved 09/14/2020 from http://journals.sagepub.com/doi/10.1177/154193129804202110.

[44] Afsaneh Koohestani, Darius Nahavandi, Houshyar Asadi, Parham M. Kebria, Abbas Khosravi, Roohallah Alizadehsani, and Saeid Nahavandi. 2019. A Knowledge Discovery in Motion Sickness: A Comprehensive Literature Review. *IEEE Access*, 7, 85755–85770. Retrieved 09/12/2020 from https://ieeexplore.ieee.org/document/8736744/.

[45] Frank L. Kooi and Marcel Mosch. 2006. Peripheral Motion Displays: Tapping the Potential of the Visual Periphery. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50, 16, (October 2006), 1604–1608. Retrieved 03/04/2018 from http://journals.sagepub.com/doi/10.1177/154193120605001619.

[46] W. A. Kukull and M. Ganguli. 2012. Generalizability: The trees, the forest, and the low-hanging fruit. *Neurology*, 78, 23, (June 2012), 1886–1891. Retrieved 09/17/2020 from http://www.neurology.org/cgi/doi/10.1212/WNL.0b013e318258f812.

[47] Joseph J. LaViola. 2000. A discussion of cybersickness in virtual environments. *ACM SIGCHI Bulletin*, 32, 1, (January 2000), 47–56. Retrieved 09/12/2020 from https://dl.acm.org/doi/10.1145/333329.333344.

[48] Hyun Kyoon Lim, Kyungha Ji, Ye Shin Woo, Dong-uk Han, Dong-Hyun Lee, Sun Gu Nam, and Kyoung-Mi Jang. 2021. Test-retest reliability of the virtual reality sickness evaluation using electroencephalography (EEG). *Neuroscience Letters*, 743, 135589. http://www.sciencedirect.com/science/article/pii/S0304394020308594.

[49] J.J.-W. Lin, H.B.L. Duh, D.E. Parker, H. Abi-Rached, and T.A. Furness. 2002. Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment. In IEEE Comput. Soc, 164–171. Retrieved 03/04/2018 from http://ieeexplore.ieee.org/document/996519/.

[50] Lili Liu, Benjamin Watson, and Masako Miyazaki. 1999. VR for the Elderly: Quantitative and Qualitative Differences in Performance with a Driving Simulator. *CyberPsychology & Behavior*, 2, 6, (December 1999), 567–576. Retrieved 09/14/2020 from http://www.liebertpub.com/doi/10.1089/cpb.1999.2.567.

[51] Roman Luks and Fotis Liarokapis. 2019. Investigating motion sickness techniques for immersive virtual environments. In *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments*. ACM, Rhodes Greece, (June 2019), 280–288. Retrieved 09/15/2020 from https://dl.acm.org/doi/10.1145/3316782.3321535.

[52] Sergo Martirosov and Pavel Kopecek. 2017. Cyber Sickness in Virtual Reality - Literature Review. In *DAAAM Proceedings*. (1st edition). Volume 1. Branko Katalinic, editor. DAAAM International Vienna, 0718–0726. Retrieved 09/12/2020 from http://www.daaam.info/Downloads/Pdfs/proceedings/proceedings_2017/101.pdf.

[53] Nicole A. Matas, Ted Nettelbeck, and Nicholas R. Burns. 2015. Dropout during a driving simulator study: A survival analysis. *Journal of Safety Research*, 55, (December 2015), 159–169. Retrieved 09/16/2020 from https://linkinghub.elsevier.com/retrieve/pii/S0022437515000730.

[54] Mark McGill, Daniel Boland, Roderick Murray-Smith, and Stephen Brewster. 2015. A Dose of Reality: Overcoming Usability Challenges in VR Head-Mounted Displays. In ACM Press, 2143–2152. Retrieved 03/18/2018 from http://dl.acm.org/citation.cfm?doid=2702123.2702382.

[55] Justin Maximilian Mittelstaedt, Jan Wacker, and Dirk Stelling. 2019. VR aftereffect and the relation of cybersickness and cognitive performance. *Virtual Reality*, 23, 2, (June 2019), 143–154. Retrieved 09/15/2020 from http://link.springer.com/10.1007/s10055-018-0370-3.

[56] D. Moher, A. Liberati, J. Tetzlaff, D. G Altman, and for the PRISMA Group. 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*, 339, jul21 1, (July 2009), b2535–b2535. Retrieved 09/13/2020 from http://www.bmj.com/cgi/doi/10.1136/bmj.b2535.

[57] Jantsje M. Mol. 2019. Goggles in the lab: Economic experiments in immersive virtual environments. *Journal of Behavioral and Experimental Economics*, 79, (April 2019), 155–164. Retrieved 09/12/2020 from https://linkinghub.elsevier.com/retrieve/pii/S2214804319300801.

[58] Jason D. Moss and Eric R. Muth. 2011. Characteristics of head-mounted displays and their effects on simulator sickness. *Human Factors*, 53, 3, 308–319. PMID: 21830515. eprint: https://doi.org/10.1177/0018720811405196. https://doi.org/10.1177/0018720811405196.

[59] Justin Munafo, Meg Diedrick, and Thomas A. Stoffregen. 2017. The virtual reality head-mounted display Oculus Rift induces motion sickness and is sexist in its effects. *Experimental Brain Research*, 235, 3, (March 2017), 889–901. Retrieved 09/08/2020 from http://link.springer.com/10.1007/s00221-016-4846-7.

[60] Benson G. Munyan, Sandra M. Neer, Deborah C. Beidel, and Florian Jentsch. 2016. Olfactory Stimuli Increase Presence in Virtual Environments. *PLOS ONE*, 11, 6, (June 2016), e0157568. Cristoforo Scavone, editor. Retrieved 09/15/2020 from https://dx.plos.org/10.1371/journal.pone.0157568.

[61] E. R. Muth, R. M. Stern, J. F. Thayer, and K. L. Koch. 1996. Assessment of the multiple dimensions of nausea: the Nausea Profile (NP). *Journal of Psychosomatic Research*, 40, 5, (May 1996), 511–520.

[62] Solène Neyret, Xavi Navarro, Alejandro Beacco, Ramon Oliva, Pierre Bourdin, Jose Valenzuela, Itxaso Barberia, and Mel Slater. 2020. An Embodied Perspective as a Victim of Sexual Harassment in Virtual Reality Reduces Action Conformity in a Later Milgram Obedience Scenario. *Scientific Reports*, 10, 1, (December 2020), 6207. Retrieved 09/17/2020 from http://www.nature.com/articles/s41598-020-62932-w.

[63] Anh Nguyen, Yannick Rothacher, Bigna Lenggenhager, Peter Brugger, and Andreas Kunz. 2018. Individual differences and impact of gender on curvature redirection thresholds. In *Proceedings of the 15th ACM Symposium on Applied Perception*. ACM, Vancouver British Columbia Canada, (August 2018), 1–4. Retrieved 09/15/2020 from https://dl.acm.org/doi/10.1145/3225153.3225155.

[64] Sarah Nichols. 2000. Individual Characteristics and Experiences of Virtual Reality Induced Symptoms and Effects (Vrise). *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 44, 5, (July 2000), 538–541. Retrieved 09/14/2020 from http://journals.sagepub.com/doi/10.1177/154193120004400514.

[65] Dilay Seda Özgen, Yasemin Afacan, and Elif Sürer. 2019. Usability of virtual reality for basic design education: a comparative study with paper-based design. *International Journal of Technology and Design Education*, (November 2019). Retrieved 09/17/2020 from http://link.springer.com/10.1007/s10798-019-09554-0.

[66] A. H. Park and S. Hu. 1999. Gender differences in motion sickness history and susceptibility to optokinetic rotation-induced motion sickness. *Aviation, Space, and Environmental Medicine*, 70, 11, (November 1999), 1077–1080.

[67] Andrew Paroz and Leigh Ellen Potter. 2017. Cybersickness and migraine triggers: exploring common ground. In *Proceedings of the 29th Australian Conference on Computer-Human Interaction - OZCHI '17*. ACM Press, Brisbane, Queensland, Australia, 417–421. Retrieved 08/23/2020 from http://dl.acm.org/citation.cfm?doid=3152771.3156148.

[68] 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, (May 2020), 1945–1954. Retrieved 09/08/2020 from https://ieeexplore.ieee.org/document/8998141/.

[69] Roos Pot-Kolder, Wim Veling, Jacqueline Counotte, and Mark van der Gaag. 2018. Anxiety Partially Mediates Cybersickness Symptoms in Immersive Virtual Reality Environments. *Cyberpsychology, Behavior, and Social Networking*, 21, 3, (March 2018), 187–193. Retrieved 09/15/2020 from http://www.liebertpub.com/doi/10.1089/cyber.2017.0082.

[70] J. T. Reason. 1978. Motion sickness adaptation: a neural mismatch model. *Journal of the Royal Society of Medicine*, 71, 11, (November 1978), 819–829.

[71] Jonas Reichenberger, Michael Pfaller, Diana Forster, Jennifer Gerczuk, Youssef Shiban, and Andreas Mühlberger. 2019. Men Scare Me More: Gender Differences in Social Fear Conditioning in Virtual Reality. *Frontiers in Psychology*, 10.

[72] Christopher J. Rich and Curt C. Braun. 1996. Assessing the Impact of Control and Sensory Compatibility on Sickness in Virtual Environments. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 40, 22, (October 1996), 1122–1125. Retrieved 09/16/2020 from http://journals.sagepub.com/doi/10.1177/154193129604002203.

[73] Dimitrios Saredakis, Ancret Szpak, Brandon Birckhead, Hannah A. D. Keage, Albert Rizzo, and Tobias Loetscher. 2020. Factors Associated With Virtual Reality Sickness in Head-Mounted Displays: A Systematic Review and Meta-Analysis. *Frontiers in Human Neuroscience*, 14, (March 2020), 96. Retrieved 09/14/2020 from https://www.frontiersin.org/article/10.3389/fnhum.2020.00096/full.

[74] Morgan Klaus Scheuerman, Katta Spiel, Oliver L. Haimson, Foad Hamidi, and Stacy M. Branham. 2020. HCI Guidelines for Gender Equity and Inclusivity (Version 1.1). (May 2020). Retrieved 09/16/2020 from HCI%20Guidelines%20for%20Gender%20Equity%20and%20Inclusivity.

[75] Jonas Schild, Joseph LaViola, and Maic Masuch. 2012. Understanding user experience in stereoscopic 3D games. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*. ACM Press, Austin, Texas, USA, 89. Retrieved 09/14/2020 from http://dl.acm.org/citation.cfm?doid=2207676.2207690.

[76] Susan M. Schneider, Mathew Ellis, William T. Coombs, Erin L. Shonkwiler, and Linda C. Folsom. 2003. Virtual Reality Intervention for Older Women with Breast Cancer. *CyberPsychology & Behavior*, 6, 3, (June 2003), 301–307. Retrieved 09/14/2020 from http://www.liebertpub.com/doi/10.1089/109493103322011605.

[77] Valentin Schwind, Pascal Knierim, Cagri Tasci, Patrick Franczak, Nico Haas, and Niels Henze. 2017. "These are not my hands!": Effect of Gender on the Perception of Avatar Hands in Virtual Reality. In ACM Press, 1577–1582. Retrieved 03/18/2018 from http://dl.acm.org/citation.cfm?doid=3025453.3025602.

[78] Daniel Shafer, Corey Carbonara, and Michael Korpi. 2017. Modern Virtual Reality Technology: Cybersickness, Sense of Presence, and Gender. *Media Psychology Review*, 11, 2. http://mprcenter.org/review/modern-virtual-reality-technology-cybersickness-sense-of-presence-and-gender/.

[79] L. James Smart, Thomas A. Stoffregen, and Benoît G. Bardy. 2002. Visually Induced Motion Sickness Predicted by Postural Instability. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 44, 3, (September 2002), 451–465. Retrieved 08/13/2020 from http://journals.sagepub.com/doi/10.1518/0018720024497745.

[80] Katta Spiel, Oliver L. Haimson, and Danielle Lottridge. 2019. How to do better with gender on surveys: a guide for hci researchers. *Interactions*, 26, 4, (June 2019), 62–65. https://doi-org.proxy.lib.uwaterloo.ca/10.1145/3338283.

[81] Kay Stanney, Cali Fidopiastis, and Linda Foster. 2020. Virtual Reality Is Sexist: But It Does Not Have to Be. *Frontiers in Robotics and AI*, 7, (January 2020), 4. Retrieved 09/17/2020 from https://www.frontiersin.org/article/10.3389/frobt.2020.00004/full.

[82] Kay M Stanney, Robert S Kennedy, and Robert Breaux. 1999. Virtual Environment Exposure Drop-Out Thresholds. *rd ANNUAL MEETING*, 5.

[83] Kay M. Stanney, Kelly S. Hale, Isabelina Nahmens, and Robert S. Kennedy. 2003. What to expect from immersive virtual environment exposure: influences of gender, body mass index, and past experience. *Human Factors*, 45, 3, 504–520. PMID: 14702999. eprint: https://doi.org/10.1518/hfes.45.3.504.27254. https://doi.org/10.1518/hfes.45.3.504.27254.

[84] Kay M. Stanney, Robert S. Kennedy, and Julie M. Drexler. 1997. Cybersickness is Not Simulator Sickness. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 41, 2, (October 1997), 1138–1142. Retrieved 09/09/2020 from http://journals.sagepub.com/doi/10.1177/107118139704100292.

[85] Kay M. Stanney, Kelly S. Kingdon, and Robert S. Kennedy. 2002. Dropouts and Aftereffects: Examining General Accessibility to Virtual Environment Technology. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 46, 26, (September 2002), 2114–2118. Retrieved 09/17/2020 from http://journals.sagepub.com/doi/10.1177/154193120204602603.

[86] Jonathan Stevens, Peter Kincaid, and Robert Sottilare. 2015. Visual modality research in virtual and mixed reality simulation. *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology*, 12, 4, (October 2015), 519–537. Retrieved 09/12/2020 from http://journals.sagepub.com/doi/10.1177/1548512915569742.

[87] William B. Stone III. 2017. *Psychometric evaluation of the Simulator Sickness Questionnaire as a measure of cybersickness*. Doctor of Philosophy. Iowa State University, Digital Repository, Ames. Retrieved 01/06/2021 from https://lib.dr.iastate.edu/etd/15429/. Pages: 11054762.

[88] A. G. Sutcliffe and K. Deol Kaur. 2000. Evaluating the usability of virtual reality user interfaces. *Behaviour & Information Technology*, 19, 6, (January

2000), 415–426. Retrieved 09/17/2020 from http://www.tandfonline.com/doi/abs/10.1080/014492900750052679.

[89] Desney S. Tan, Darren Gergle, Peter G. Scupelli, and Randy Pausch. 2004. Physically large displays improve path integration in 3D virtual navigation tasks. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 439–446.

[90] Nana Tian, Romain Clément, Phil Lopes, and Ronan Boulic. 2020. On the effect of the vertical axis alignment on cybersickness and game experience in a supine posture. In *2020 IEEE Conference on Games (CoG)*, 359–366.

[91] Ryan Tyrrell, Hilla Sarig-Bahat, Katrina Williams, Grace Williams, and Julia Treleaven. 2018. Simulator sickness in patients with neck pain and vestibular pathology during virtual reality tasks. *Virtual Reality*, 22, 3, (September 2018), 211–219. Retrieved 09/14/2020 from http://link.springer.com/10.1007/s10055-017-0324-1.

[92] Martin Usoh, Kevin Arthur, Mary C. Whitton, Rui Bastos, Anthony Steed, Mel Slater, and Frederick P. Brooks. 1999. Walking > walking-in-place > flying, in virtual environments. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques - SIGGRAPH '99*. ACM Press, Not Known, 359–364. Retrieved 09/17/2020 from http://portal.acm.org/citation.cfm?doid=311535.311589.

[93] Séamas Weech, Jessy Parokaran Varghese, and Michael Barnett-Cowan. 2018. Estimating the sensorimotor components of cybersickness. *Journal of Neurophysiology*, 120, 5, (November 2018), 2201–2217.

[94] Michael L. Wilson and Amelia J. Kinsela. 2017. Absence of Gender Differences in Actual Induced HMD Motion Sickness vs. Pretrial Susceptibility Ratings. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 61, 1, (September 2017), 1313–1316. Retrieved 08/14/2020 from http://journals.sagepub.com/doi/10.1177/1541931213601810.

[95] Bob G. Witmer and Michael J. Singer. 1998. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments*, 7, 3, (June 1998), 225–240. Retrieved 03/21/2018 from http://www.mitpressjournals.org/doi/10.1162/105474698565686.

[96] Matthew Wood, Gavin Wood, and Madeline Balaam. 2017. "They're Just Tixel Pits, Man": Disputing the 'Reality' of Virtual Reality Pornography through the Story Completion Method. In ACM Press, 5439–5451. Retrieved 03/18/2018 from http://dl.acm.org/citation.cfm?doid=3025453.3025762.

[97] Robert Xiao and Hrvoje Benko. 2016. Augmenting the Field-of-View of Head-Mounted Displays with Sparse Peripheral Displays. In ACM Press, 1221–1232. Retrieved 09/27/2016 from http://dl.acm.org/citation.cfm?doid=2858036.2858212.

[98] Hiroo Yamamura, Holger Baldauf, and Kai Kunze. 2020. Pleasant locomotion – towards reducing cybersickness using fnirs during walking events in vr. In *Adjunct Publication of the 33rd Annual ACM Symposium on User Interface Software and Technology* (UIST '20 Adjunct). Association for Computing Machinery, Virtual Event, USA, 56–58. https://doi.org/10.1145/3379350.3416184.

[99] P. Zimmons and A. Panter. 2003. The influence of rendering quality on presence and task performance in a virtual environment. In *IEEE Virtual Reality, 2003. Proceedings.* IEEE Comput. Soc, Los Angeles, CA, USA, 293–294. Retrieved 09/15/2020 from http://ieeexplore.ieee.org/document/1191170/.