Remote VR Studies

A Framework for Running Virtual Reality Studies Remotely Via Participant-Owned HMDs

RIVU RADIAH, Bundeswehr University Munich, Germany

VILLE MÄKELÄ, University of Waterloo, Canada and LMU Munich, Germany

SARAH PRANGE, Bundeswehr University Munich, Germany

SARAH DELGADO RODRIGUEZ, Bundeswehr University Munich, Germany

ROBIN PIENING, LMU Munich, Germany

YUMENG ZHOU, LMU Munich, Germany

KAY KÖHLE, LMU Munich, Germany

KEN PFEUFFER, Bundeswehr University Munich, Germany

YOMNA ABDELRAHMAN, Bundeswehr University Munich, Germany

MATTHIAS HOPPE, LMU Munich, Germany

ALBRECHT SCHMIDT, LMU Munich, Germany

FLORIAN ALT, Bundeswehr University Munich, Germany

We investigate opportunities and challenges of running virtual reality (VR) studies remotely. Today, many consumers own head-mounted displays (HMDs), allowing them to participate in scientific studies from their homes using their own equipment. Researchers can benefit from this approach by being able to recruit study populations normally out of their reach, and to conduct research at times when it is difficult to get people into the lab (cf. the COVID pandemic). In an initial online survey (N=227), we assessed HMD owners' demographics, their VR setups and their attitudes towards remote participation. We then identified different approaches to running remote studies and conducted two case studies for an in-depth understanding. We synthesize our findings into a framework for remote VR studies, discuss strengths and weaknesses of the different approaches, and derive best practices. Our work is valuable for HCI researchers conducting VR studies outside labs.

CCS Concepts: • **Human-centered computing** → **User studies**; **Virtual reality**.

Additional Key Words and Phrases: Virtual Reality, User Studies, Remote Studies, Data Collection Methods

Authors' addresses: Rivu Radiah, sheikh.rivu@unibw.de, Bundeswehr University Munich, Germany; Ville Mäkelä, vmakela@uwaterloo.ca, University of Waterloo, Canada, LMU Munich, Germany; Sarah Prange, sarah.prange@unibw.de, Bundeswehr University Munich, Germany; Sarah Delgado Rodriguez, sarah.delgado@unibw.de, Bundeswehr University Munich, Munich, Germany; Robin Piening, robin.piening@unibw.de, LMU Munich, Germany; Yumeng Zhou, LMU Munich, Germany; Kay Köhle, LMU Munich, Germany; Ken Pfeuffer, ken.pfeuffer@unibw.de, Bundeswehr University Munich, Munich, Germany; Yomna Abdelrahman, yomna.abdelrahman@unibw.de, Bundeswehr University Munich, Germany; Matthias Hoppe, matthias.hoppe@ifi.lmu.de, LMU Munich, Germany; Albrecht Schmidt, albrecht.schmidt@ifi.lmu.de, LMU Munich, Germany; Florian Alt, florian.alt@unibw.de, Bundeswehr University Munich, Germany.

1 INTRODUCTION

Virtual Reality (VR) has become a widely adopted technology in the Human-Computer Interaction (HCI) research community and beyond [64]. Many subtopics emerged around this technology, including but not limited to, novel interaction techniques, presence and immersion, avatar modeling, navigation, and locomotion. More recently, researchers started looking into how VR can substitute or complement other research approaches [41, 59], for example, in-the-wild studies that require a lot of effort to maintain the technology used, evaluations in potentially dangerous environments (e.g., automotive or pedestrian user interfaces), or the investigation of situations that occur rarely.

In the past decades, VR headsets were not widely available and their setup required a substantial degree of technical knowledge. Hence, the vast majority of research was conducted in lab settings, where researchers took care of the setup and guided participants through (controlled) studies. As a result, participants of such studies were often part of a rather homogeneous university population – a well-known challenge in HCI research [8].

The past years witnessed a proliferation of VR head-mounted displays (HMDs) in the consumer market[1] and while the number of worldwide shipped VR units with 6 million units in 2019[2] is still marginal from a commercial perspective, a considerable number of end users now own a VR headset. This creates an unprecedented chance for researchers to move VR research out of the lab and reach a more diverse community – similar to what happened with other technologies that reached the consumer market, such as smartphones [26–28] or public displays [18].

In this work, we contribute an in-depth investigation of an emerging research paradigm, that is conducting out-of-the-lab research with owners of VR headsets. In particular, we explore the challenges and pitfalls researchers are facing as they are shifting or complementing their research using this new paradigm and synthesize them into a framework for use by other researchers. To this end, our work is guided by the following research questions:

- **RQ1:** How suitable are HMD owners and their VR setups as subjects for remote studies, and what limitations do home setups have?
  - **1.1:** What are the demographics of HMD owners?
  - **1.2:** Which VR equipment do they own?
  - **1.3:** What are their home VR setups like?
  - **1.4:** How do they feel about participating in remote studies from home?
- **RQ2:** What suitable approaches exist for conducting remote VR studies, and what are their advantages and disadvantages?
  - **2.1:** What options exist for developing and setting up VR research applications?
  - **2.2:** How can VR research prototypes be distributed to study participants?
  - **2.3:** Through which channels can owners of VR headsets be reached and recruited?
  - **2.4:** What special considerations are there when designing and running a remote study?

Our research approach is as follows: we set out with conducting an online survey (N=227) among owners of VR headsets to understand their demographics, the equipment they own, the setting in which they use it, and whether they would be willing to participate in remote VR studies. Subsequently, based on a review of existing VR distribution channels, we describe different ways in which VR applications for remote studies can be implemented and distributed as well as how data collection can be realized. We then report on our experiences from two remote case studies, where we experimented with different approaches. We then synthesize all our findings into a framework that guides researchers through different ways of designing and conducting remote VR studies.

---

[1]Consumer VR market size: https://www.statista.com/statistics/528779/virtual-reality-market-size-worldwide/
[2]Shipment of VR devices by vendor: https://www.statista.com/statistics/671403/global-virtual-reality-device-shipments-by-vendor/

Our framework consists of four primary approaches to remote studies:

(1) Researchers build a standalone VR application that is distributed directly to participants.
(2) Researchers build a VR application that is distributed through existing vendor platforms ("app stores") such as Steam and the Play Store.
(3) Researchers build a VR application using an API of an existing social VR platform (e.g., Rec Room, VRChat) and upload it to the corresponding platform.
(4) Researchers set up their study environment directly within an existing social VR platform using the tools provided by these platforms.

All of these approaches come with unique advantages and drawbacks. Most notably, the approaches vary in how much they utilize and rely on existing platforms and services. The approach should be primarily chosen based on the planned user study's specific needs, for example, which interactive capabilities are needed and what data needs to be collected. Using existing vendor platforms and social VR platforms requires that researchers conform to the regulations and limitations of these platforms. This often means that options for data collection are limited, and the platforms are also more limited in their functionalities. However, using such platforms can often make setting up the study far easier and faster, and existing platforms ensure an existing user base and alleviate stability and compatibility issues. Building custom software and distributing it independently may get around many of the limitations with existing platforms, but the workload is often higher and may result in other challenges (e.g., ensuring the safety and privacy of participants).

Another dimension is whether or not the VR study is conducted with a remote study experimenter, or whether participants run the study independently, i.e., without remote guidance. We discuss how well the four main approaches lend themselves to guided studies and independent studies.

Our research is valuable for researchers considering to conduct VR studies remotely. Remote VR studies offer researchers the opportunity to study users in their natural settings (e.g., using their own VR setups) and to recruit participants otherwise out of their reach, be it due to geographic limitations, target groups that are difficult to find, or specific events. Most prominently, the recent COVID-19 pandemic affected the HCI community around the globe, making it difficult for many researchers to recruit participants for their work in the lab. This unprecedented event raised the question: "How can VR researchers continue to conduct studies?" Here, our work can provide useful guidance. We discuss the strengths and weaknesses of different remote study approaches, and identify ways to assess which approach might be the most suitable for a planned study.

Contribution Statement. We contribute a holistic view on remote VR studies. Specifically, (1) we present the results of an online survey assessing users' view towards such studies and their VR settings; (2) we provide a review of research approaches with a focus on how they influence the development / study setup, software distribution, and recruitment challenges; (3) we report on two studies and derive lessons learned; and (4) we synthesize our findings into a framework.

2 BACKGROUND

Several strands of prior work are relevant to our research. After briefly introducing the term *remote VR studies*, we review (a) work involving ubiquitous technologies for conducting out-of-the-lab research as well as (b) prior work on conducting VR studies outside the lab.

2.1 Terminology and Scope

Over many decades, the HCI research community adapted or came up with new approaches to research, which can be broadly classified into research conducted in controlled settings (e.g., in the lab) and in less controlled settings (e.g., in the field). We have seen researchers conducting VR studies in the lab as well as in the field [40]. In both cases, researchers have typically been on-site instructing study participants, overseeing the procedure, and collecting data.

In many areas of HCI, research has also been conducted remotely, in such a way that researchers and participants are not co-located. We believe that specifically work in VR can benefit from this approach, as people can participate from their *homes using their own head-mounted displays*. In this paper, we refer to these type of studies as *remote VR studies*.

Head-mounted displays can also differ considerably regarding their capabilities, ranging from low-cost solutions (e.g., Google Cardboard[3]) to high-end devices (e.g., HTC Vive Pro and similar HMDs with inside-out or outside-in tracking and hand-held controllers). While we intend our work to be relevant for all kinds of HMDs, we focus more on high-end HMDs that use tracking stations and controllers, as these setups generally require more effort during their set up and might pose unforeseen challenges when leveraged for remote studies.

2.2   Leveraging Consumer Devices for Out-of-the-Lab Research

The idea of leveraging consumers' devices for 'bringing the lab to participants' is already established practice in several areas of human-computer interaction research. Among the most prominent examples is work by Henze et al. who used the consumers' smartphones to conduct out-of-the-lab studies, initially on touch targeting and typing behavior [26–28]. Over the following years, the community saw more studies focusing on various aspects, including but not limited to an exploration of user interaction with notifications [52], authentication behavior [5, 12, 53], and keystroke dynamics [11]. Apart from smartphones, researchers have also leveraged other personal devices of participants, such as smartwatches [38].

Prior work has focused on understanding both strength and challenges of various types of studies. Of particular interest is the work of Gustarini et al. [25] who categorize challenges into the design, development, execution, and data analysis phase. While their work is specific to studies focusing on smart phones, several of the challenges apply to other technologies as well and, hence, informed our work. These challenges include, among others, which data can be collected, where data can be stored, how researchers can deal with device heterogeneity, recruiting bias, remunerating participants, participant cheating, handling participants' questions, considering participants' motivation and privacy needs, and conducting interviews. Throughout this work, we touch on these challenges, particularly from the perspective of remote VR studies.

Studies have also been conducted to understand the extent of ecological validity in remote experiments. For example, Andreasen et al. [6] conducted an empirical comparison of remote usability testing and conventional lab testing, and Germine et al. [23] investigated data quality across lab-based and web-based experiments. Research shows that it is possible to achieve ecologically valid results using out-of-lab studies [6, 23].

2.3   Out-of-the-Lab VR Studies

In the following section we summarize key insights obtained from prior research. We take a chronological view on how out-of-the-lab studies have progressed over the years.

In 2012, Hodgson el at. [31] presented ways for 'portable' VR studies, using systems that can be easily carried and used in-the-wild. The authors introduced a self-contained backpack, capable of running VR simulations and demonstrated how such portable systems can be used by researchers.

Four years later, Steed et al. [58] experimented on presence and embodiment in VR through remote studies by recruiting Samsung Gear VR and Google Cardboard owners. They argue that for remote VR studies it is generally easier to recruit more participants than for typical lab-based studies. However, their work suggests that the diversity of participants is not necessarily inherent

---

[3]In Cardboard VR, users insert their smartphone into a cardboard frame to build a simple VR headset, c.f. https://arvr.google.com/cardboard

in the remote approach, but is still specific to the study. The authors also highlighted that remote studies might require more effort than lab studies in terms of, for example, preparation and app development. Bramley et al. [10], on the other hand, sent out Google Cardboards to a large amount of participants to be paired with the participants' own smartphones, so that they could experience a VR film before filling in an online mobile survey.

In 2017, Mottelson et al. [46] conducted experiments both in and out of the lab to understand the advantages and pitfalls of both approaches. The experiments generally yielded comparable results as both in-lab and remote studies were able to collect reliable data and similar output. At the same time, out-of-the-lab experiments were characterized by a greater heterogeneity among participants such as sample population. They also outline several limitations of remote studies including demographics validity, ethical concerns, and the lack of experiment control.

Ma et al. [39] conducted three behavioral experiments in VR using crowd sourcing in 2018. Their evaluation suggests the feasibility of conducting web-based VR experiments albeit several challenges, including few participants with access to VR-capable devices. The authors also highlight many advantages, including obtaining a more diverse sample compared to lab VR studies. In a study about social interactions in weekly VR get-togethers, Moustafa and Steed [47] also highlighted the potential of reaching broad and diverse audiences through remote studies. VR has also been studied in educational contexts, for example, Makransky et al. [43] found that their desktop VR experience resulted in comparable learning outcomes both in the classroom and at home.

Recently, Saffo et al. investigated the feasibility of running VR studies via a popular social platform, VRChat [51]. They identified the ability to enable researchers to create their own content, the large user base, and the ease of recruiting participants from within VRChat as advantages of the platform as it is easier to recruit participants within the platform. At the same time, Saffo et al. reported on data collection to be limited when using online platforms as each online platform reviews the content before researchers are able to upload content.

Huber and Gajos [32] worked on conducting online VR studies with uncompensated and un-supervised participants. While they identified advantages, such as recruiting participants from all over the world with various backgrounds, there are challenges: for example, in this study the sample count was much smaller compared to other online studies.

2.4 Summary

Our review shows that while researchers have recognized and acknowledged the potential of remote VR studies, for many years they have struggled with challenges, such as the available technology, the number of people who had access to the technology, means to reach out to them, and ways of delivering software to the participants. Not only with the current pandemic, we see that interest in this type of studies is increasing at a rapidly accelerating pace. We believe to be currently at a turning point where with advances in technology and due to the increasing proliferation of the technology for the first time it is finally possible for VR researchers to strongly benefit from this approach. Yet there is still a lack of a comprehensive understanding of how such studies can be implemented. This is at the focus of our work.

While much can be learned from other sub fields of HCI where technology that became widely available to end users was leveraged for conducting remote studies, there are many aspects unique to VR and head-mounted displays that remain under-explored, including but not limited to the available technology, the environment, data collection methods, and distribution channels.

Some of the aforementioned aspects have been touched upon in prior work. Yet, our analysis shows that pre-requisites have changed over the years and new opportunities and challenges have emerged. Our work provides an understanding of the current state-of-the-art in running remote VR studies as well as a comprehensive assessment of this study paradigm that we synthesize

into a framework. The framework explains different approaches to remote VR studies and also summarizes best practices and lessons learned. In doing so, we hope to offer useful guidance to researchers conducting VR studies, similar to what prior research has offered in other areas of HCI (e.g., public displays [4, 42, 44].)

3 UNDERSTANDING VR USERS THROUGH AN ONLINE SURVEY

We conducted an online survey among VR users with the goal of better understanding (a) their demographics, (b) how they can be reached and recruited, (c) what their VR setups look like, (d) which technology they own, and (e) whether they are willing to participate in remote VR studies.

    We used SoSci, a survey tool installed on our institution's web server. This way, data was neither handled nor stored by any external parties, making data collection GDPR-compliant [2, 24].

3.1 Survey Content and Questions

In particular, our survey focuses on the following aspects:

> **Demographics** We assessed the participants' age, gender, background, and their reasons for owning a VR device. Our intention here was to gain a preliminary understanding of the diversity of HMD owners. In many other contexts, past research highlights the diversity of the study population as one of the primary strengths of remote studies [26]. Similar arguments have been made for individual remote VR studies [39, 46], but we lack a broad understanding of the demographic of HMD owners.

> **VR Equipment & Platforms** We asked about the devices they own and the platforms they are using. Device heterogeneity is one of the challenges in remote studies [25] and may lead to technical issues [27]. Additionally, because we envision that social VR platforms can be used to conduct some types of remote VR studies, we need to investigate which platforms are being commonly used, and whether users focus on one or several platforms.

> **VR Setting and Use** We asked them to describe their VR setup, the space, how often they use it, and how likely they are to be interrupted while being in VR. A lot of VR research relies on physical movement, which in turn is dependent on the available tracking space [41]. We are not aware of any existing research that has investigated the size and nature of HMD owners' tracking spaces at home. The amount of space affects the feasibility of the planned study in home environments, and also ethical considerations, such that the remote participants are not put at risk if their home setups are too small for the intended interactions. Finally, users might be interrupted often in more natural and less controlled contexts, such as when using smartphones [27]. We hypothesize that interruptions might be particularly critical in VR studies due to being enclosed in the virtual world and using the HMD, as taking off the HMD is disruptive and can break the sense of presence [56].

> **Willingness to participate** We asked the participants how willing they are to participate in remote VR studies, both in general as well as in situations such as the COVID-19 lockdown, and what kind of payment they might expect. To the best of our knowledge, prior research has not investigated how people might perceive remote study participation, and whether they would feel comfortable participating in VR studies remotely. We return to this discussion in our case studies, where we asked participants' about their experience of participating in a remote VR study.

3.2 Recruiting

We distributed the survey via different channels, in particular such that would also be used by researchers to reach out to potential participants of a VR study. To this end, we included research-oriented crowd sourcing platforms, VR online forums, social media and university mailing lists. In the following we report some lessons learned.

3.2.1 Research-oriented Crowd Sourcing Platforms. We recruited through two platforms: Prolific[4] and XRDRN[5]. Prolific is a platform primarily meant for recruiting study participants, for example, for surveys. It allows participants to be pre-selected by demographics (for example, owners of VR devices). XRDRN is a dedicated platform to connect researchers and participants with a specific focus on mixed reality. The platform was built specifically in response to the COVID-19 pandemic. Response rates from Prolific were considerably higher and faster. This was likely because (a) Prolific integrates means for payment directly with the platform whereas for XRDRN this would be handled out of the platform between researchers and participants, and (b) XRDRN is still in the early stages of building a community. Another aspect is that since all participants on Prolific are paid, many of them can be expected to be 'professional' study takers.

3.2.2 Forums & Social Media. We reached out to VR users through Facebook, Reddit and platform-specific forums (e.g., Steam, Rec Room). Within these platforms, we identified suitable subgroups (e.g., VR-related subreddits in Reddit).

Within platforms, subgroups could be easily identified. However, it took some time to ensure that we complied with the rules of each subgroup and that posting an invitation to an online survey was appropriate. Where groups did not specifically allow or disallow for these kinds of posts, we reached out to moderators to seek permission. While for some groups we obtained permission in this way, many moderators were unsure and suggested we rather post in other channels, for example, such that were concerned with the organization of VR events or for discussing miscellaneous topics. We noticed that such groups often had fewer active members, which limited the visibility of the survey.

Some subgroups employed specific posting procedures. Some required specific tags or features to be assigned to messages (for example, on Reddit or Rec Room) – yet the pre-defined tags were often a poor match for our recruitment messages. Many forums also make it mandatory for users to achieve a certain number of points (points are acquired based on several metrics such as number of comments and posts on other discussion threads) before they are eligible to create their own threads, while others required a user account to be at least two months old before being eligible to post. When posting links to different subforums simultaneously, there is a chance that auto-moderators flag the posts as spam. Established VR products (such as Unity and Rec Room) run forums and groups on different social platforms (such as Reddit, Facebook, Twitter, Discord). Those could be used in addition to more general groups on VR.

3.2.3 University Mailing Lists. This method helped to reach out to a large number of recipients. In particular, we used mailing lists allowing a general university population to be reached. This population included many non-technical subjects. Our experience shows that only few of them were eligible for the survey (as they were required to own an HMD to participate). Mailing lists targeting students from computer science subjects might reach more participants.

3.3 Results

We received 227 complete surveys (out of 276). Respondents were recruited through Prolific (97), Reddit (95), mailing lists (14), Facebook (12), Discord (4), VR platforms (3) and other (2).

---

[4]Prolific: https://www.prolific.co/, last accessed May 6, 2021
[5]XRDRN: https://www.xrdrn.org/, last accessed May 6, 2021

Table 1. Online survey: participants' demographics including their relationship to VR, which VR platforms they use (left), their purposes of using VR, and possible new uses discovered during the pandemic (multiple choice questions each) (right).

Gender		
	Male	170
	Female	52
	Prefer not to say	1
	Other	4
Roles		
	I am a consumer of VR products and applications.	214
	I am a developer of VR products and applications	21
	I create content for VR.	20
	I conduct research in the area of VR.	16
	Other	3
Used VR Platforms		
	VRChat	91
	None	77
	Rec Room	57
	AltspaceVR	26
	Mozilla Hubs	23
	Other	12
	Bigscreen VR	8
	Hologate VR	7
	Cluster	5
	Playstation VR	5
	The Wild	4
	A Township Tale	2

Reasons for Using VR		
	Gaming	206
	Social platforms	63
	Socializing with friends and family	43
	Development or Research	33
	Work (e.g meetings, presentation)	17
	Other	14
	Watch Videos (e.g. Netflix or films)	10
Uses Discovered during Pandemic		
	I discovered new VR games.	130
	I discovered new virtual social platforms.	35
	I started VR meetings with friends and relatives.	24
	I started watching VR movies together with friends.	23
	My usage did not change.	18
	I started VR collaboration with colleagues.	12
	Other	11

3.3.1 Demographics. A total of 227 participants responded to our online survey, 52 female (cf. Table 1 for an overview). Participants were between 18 and 56 years of age ($Mean = 29.72, SD = 8.8$). They mainly lived in the US ($N = 66$), the UK ($N = 54$), and Germany ($N = 24$).

Most of the respondents identified as *consumers* of VR products ($N = 214$, 94%), and some were VR developers ($N = 21$, 9%), VR content creators ($N = 20$, 9%), and VR researchers ($N = 16$, 7%). They mainly used their VR setup for *gaming* ($N = 206$, 91%), but also for other purposes like socialising, development and research, and during work.

We furthermore asked the participants if they had discovered new uses for their VR setup during the COVID-19 pandemic. More than half ($N = 130$, 57%) reported that they had discovered new VR games to play, and some had discovered new VR platforms ($N = 35$, 15%). Some users had started using their setups for entirely new purposes, like meeting with friends and relatives ($N = 24$, 11%), watching movies socially ($N = 23$, 10%), and collaborating with colleagues ($N = 12$, 5%).

3.3.2 Understanding Users' VR Settings. We first asked participants about the *devices they owned.* Most participants used Oculus devices ($N = 79$), followed by Valve Index ($N = 59$), PlayStation VR ($N = 41$), and HTC Vive ($N = 23$). Only a handful owned VR headsets made of cardboard.

Most participants were the sole users of their VR setup ($N = 153$, 67%). Their setups were placed in rooms rarely used by others ($Median = 2$, 1=not at all, 5=almost all the time). Consequently, most participants could use their VR setups without any notable constraints ($N = 198$), as opposed to, for example, not being able to access the VR setup due to someone else using the room. Most participants also evaluated that the chances of being interrupted during VR use were low ($Median = 2$, 1=very unlikely, 5=very likely). Furthermore, participants reported that getting their VR setup ready for use (Figure 2c) did not require much effort ($Median = 2$, 1=no effort at all, 5=a lot of effort).

Regarding usage times (Figure 1a), most participants reported to normally use their setups for up to two hours on weekdays (in total during an average week) ($N = 129$) as well as up to two hours
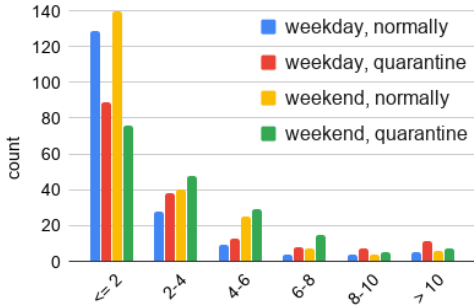
(a) Total time (hours) spent in VR on weekdays and weekends. Most users spend at most two hours with the VR setup on weekdays and on weekends under normal circumstances. During the pandemic, participants reported somewhat higher usage.
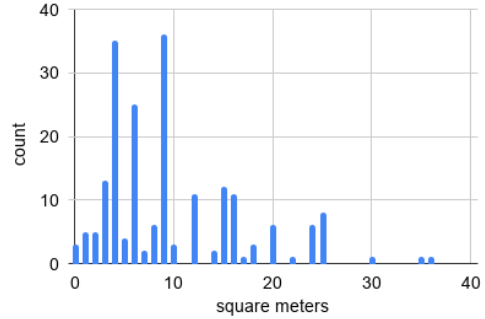
(b) Space available for users' VR setups, showing most users have a space of up to 25 square meters. 26 participants stated a larger than 40 square meters space (excluded in this plot).

Fig. 1. Results on how much time participants spend in VR and how large their VR setups are.

on weekends ($N = 140$) (in total during an average weekend). During the pandemic, participants reported spending overall more time with their VR setup.

We asked participants to additionally give an estimate of their VR setup's *size*, providing the width and length in meters (Figure 1b). 65 participants (29%) reported having up to 5 m^2, 71 (31%) from 5 to 10 m^2, 26 (11%) from 10 to 15 m^2, 21 (9%) from 15 to 20 m^2, and 15 (7%) from 20 to 25 m^2.

3.3.3 Participation in VR Studies. We inquired about participants' *willingness to participate in VR studies* ('I am willing to participate in VR studies.'; 5-Point Likert scale; 1=strongly disagree, 5=strongly agree). To better understand whether this was a result of the *current pandemic situation*, we asked the question for both during and after the lockdown (Figure 2a). 152 participants (67%) agreed or strongly agreed to be willing to participate in VR studies after the pandemic (*Median* = 4), while 165 participants (73%) agreed or strongly agreed during the pandemic (*Median* = 4).

We also looked into how *comfortable* participants felt about participating in VR studies in different study locations (Figure 2b). Participants felt more comfortable about the idea of participating from home than attending a lab study (home: *Median* = 5, lab: *Median* = 3).

Regarding *compensation* (multiple choices), participants were mostly willing to accept cash ($N = 189$, 83%), but also game vouchers ($N = 103$, 45%). A total of 77 participants (34%) stated that they would do so voluntarily without any compensation.

3.4 Summary and Discussion

Our survey provided insights on the current mainstream users of VR. First, the survey confirms the 'cliche' of VR users being primarily gamers (91%). Furthermore, users were predominantly male (77%) and mostly consumers (94%). Besides gaming, a rather prominent use case for VR was socializing, with 28% of participants stating they use social platforms and 19% that they use VR to meet with friends and family. We also found that participants used many different VR devices and platforms. The most popular *VR platforms* were VRChat (40%) and Rec Room (25%) followed by AltspaceVR (11%) and Mozilla Hubs (10%), indicating that there is a wide range of platforms with potential for researchers to conduct studies. Many participants used more than one platform.

Most HMD owners used their VR setups exclusively without notable constraints. Participants reported that the effort required to get their setup ready for use was low. Respondents also estimated

(a) Willingness to participate in studies

(b) Comfort in participating in studies
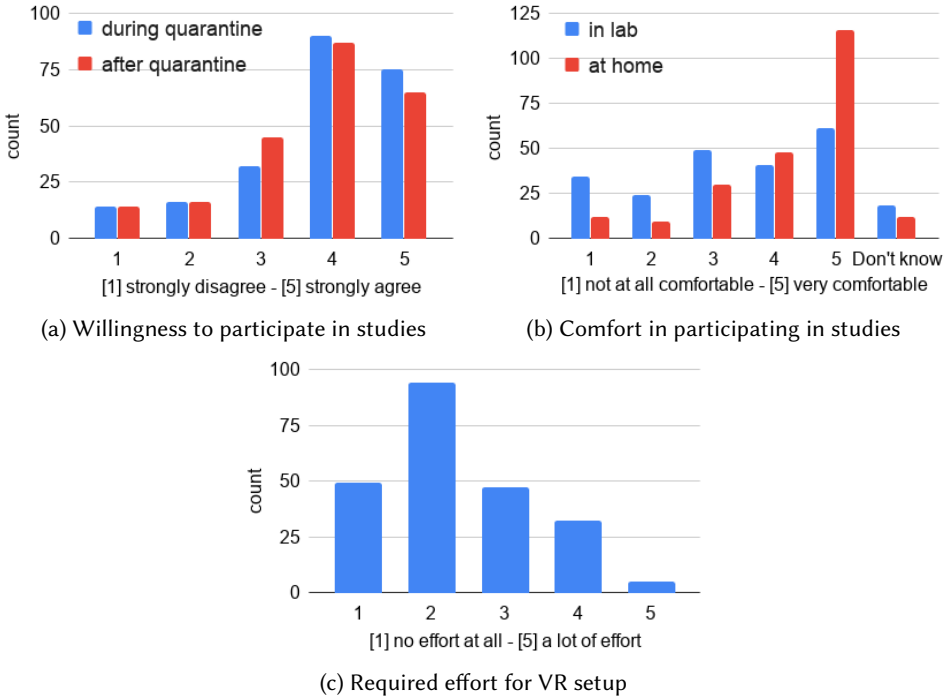

(c) Required effort for VR setup

Fig. 2. Results on how willing participants are to participate in remote studies, how comfortable they feel about it, and how much effort it is for them to use their VR setups.

that the likelihood of being interrupted during VR use was low, again suggesting that the setups are mainly used in personal spaces (e.g., bedrooms) rather than in shared spaces (e.g., living rooms). These findings indicate that home VR setups are generally well suited for remote studies.

Furthermore, we assessed the size of the users' VR setups. The majority of participants (60%) seemed to operate withing a space of up to $10\,m^2$. Around half of this group (29% of all respondents) reported having a small space of up to $5\,m^2$, likely indicating a seated setup. These findings suggest that remote studies requiring active movement in a larger space (e.g., more than $15\,m^2$) might be challenging as many HMD owners do not have the necessary space, potentially risking skewed research results as well as participants getting injured. To work around the problem of small spaces, researchers might consider various locomotion techniques, such as teleportation techniques [9, 22], to reduce the need for physical movement. However, such techniques might not always be desirable, particularly if "real" physical movement is critical to the phenomena being studied.

The clear majority of participants expressed willingness to participate in remote VR studies (67% under normal conditions and 73% during the pandemic). Participants also felt very comfortable about the idea of participating in studies from their own homes, noticeably more so than participating in lab studies. The survey attempts to assess the willingness of a potential participant to participate. In the end, participation is an amalgam of many factors and is dependent on each specific study, such as the participants actually signing up for the study and participating on it, the setup requirements and data collection methods based on the specific research question. Thus, our obtained results indicate an initial interest to participate from home. Our survey also indicates that participants are comfortable about remote participation and they have VR setups which are likely to be permanent,

private and uninterrupted during participation. We believe these initial findings are beneficial when researchers are planning remote studies.

Survey respondents also reported more VR usage as a result of the COVID-19 pandemic, and some reported having started to use their VR setups in new ways, like meeting with friends and family. This suggests that users are open to new use cases of VR. This result, together with the respondents' general willingness to participate in remote studies, further strengthens our belief that HMD owners are comfortable with participating in studies from their homes.

We distributed the survey over numerous channels to limit a potential recruiting bias. Nevertheless, we acknowledge that it is hard to fully avoid a recruitment bias. This may have impacted the survey results depending on the popularity of each channel chosen, as some VR platforms may have been more/less popular.

In conclusion, our results indicate that HMD owners are highly suitable to be recruited for remote studies. Most respondents were willing to participate in remote studies and felt comfortable about it. Their VR setups and use also support this, as most setups seem to be used exclusively with a low chance of interruptions. The most notable challenge identified through our results is that many HMD owners have relatively small setups, which might be challenging for some studies. Regarding the composition of the sample, more diverse samples can be recruited that are not typical in traditional studies (e.g., participants from different countries and cultures). At the same time, there are some traits, like the amount of VR experience and the attitudes towards VR technologies, where HMD owners may be rather homogeneous.

4 POTENTIAL APPROACHES TO REMOTE STUDIES

In this section, we discuss potential ways to conduct remote VR studies and hypothesize about their strengths and weaknesses. We also provide a brief discussion on app stores and social VR platforms that could be utilized for remote VR studies.

4.1 Development of VR Applications

4.1.1 Standalone Applications. For most VR research projects, researchers develop a custom application that runs independently. This often involves designing and modeling a digital environment, programming the desired functionality, and designing and programming the methods for data collection [41]. The advantage of custom-built applications is that they provide maximum flexibility and freedom when it comes to designing the visuals, means for interaction, data collection, and data storage, among other aspects.

Many modern tools are of great help with building VR applications. Particularly popular in this regard is the Unity engine [6], complemented with the SteamVR plugin [7]. The Unity engine is a massively popular tool for creating games, but it is also widely used for industrial purposes, films, and architecture, among others. As a result of its popularity, the internet is full of tutorials and tips for Unity developers, and communities for supporting developers are plentiful and active. The SteamVR plugin takes care of the basic communication with the head-mounted display and the VR application – many HMDs, like the HTC Vive, support SteamVR. This greatly alleviates the hurdles related to developing for HMDs. For the most part, developing VR applications is not any different from general 3D development. There are also other tools and application programming interfaces (APIs) that assist with VR development, like WebXR Device API [8] for building web-based VR experiences.

---

[6]https://unity.com/, last accessed May 6, 2021
[7]https://assetstore.unity.com/packages/tools/integration/steamvr-plugin-32647, last accessed May 6, 2021
[8]https://www.w3.org/TR/webxr/, last accessed May 6, 2021

Table 2. Comparison of social VR platforms.

Platform	Interactions	Customizable Avatars	Data Collection	External Applications and Sources	Viability
AltspaceVR	Programmable	Yes (in-game)	Full (web-hosted)	Browser	High
VRChat	Programmable	Yes (model upload)	Limited	Youtube player & browser	High
Rec Room	Partially programmable	Yes (model upload)	Limited	No	High
Mozilla Hubs	No Influence	Yes (in-game customisation or model upload)	Limited	Share media from PC or web	Medium
The Wild	No influence	No	*No information available*	Revit metadata	Low
EonReality	Partially programmable	No	*No information available*	Models and media	Low
Minecraft	Partially programmable	Yes (reskin)	No	No	Low

Despite these great tools and related resources at our disposal, it is clear that building VR applications takes time and requires considerable software engineering expertise, and in many cases, also expertise in other areas like visual design and 3D modeling. We also anticipate further challenges with remote studies, as researchers are not present in person to instruct and monitor participants. Therefore, researchers may need to develop additional tools to interact with participants or consider other ways to ensure smooth procedures.

4.1.2 Social VR Platforms. Another, emerging possibility for remote studies is to use *social VR platforms*. By social VR platforms, we refer to online platforms designed to allow users with HMDs to meet, talk, and interact with each other. Many of these platforms, like Rec Room and VRChat, offer diverse customization options that might be enough for many user studies. There are considerable differences between such VR platforms in terms of their user base and what features they provide. Based on our analysis of these platforms, we see two distinct approaches to implementing virtual environments for user studies.

First, some VR platforms allow *custom applications to be uploaded*. For example, VR applications built with the Unity engine can be uploaded to VRChat, as long as the applications use the VRChat Software Development Kit (SDK). Publishing applications in VR platforms requires that the application fulfills requirements and follows rules put forth by the platform. This may result in more restricted possibilities as opposed to fully independent VR applications. At the same time, these platforms are likely to bring various advantages. For example, installing and running the application is likely less error-prone and more novice-friendly, and compatibility of different hardware is handled by the platform rather than the application. Furthermore, it might be possible to attract study participants from the existing user base of the platform.

Second, many VR platforms allow users to *build their own environments inside the platform* using simple built-in tools, and offer customization options. This includes choosing between various indoor and outdoor settings and adding 3D objects like furniture, interactive objects like drawing boards, and even questionnaires. Although it is clear that these tools offer less flexibility and freedom for researchers than fully custom-built applications, these tools make building VR environments considerably faster and easier. Another significant advantage is that these tools are more accessible to researchers from other fields who might have less technical expertise.

We conducted an analysis (Table 2) of existing VR platforms to understand how viable they would be for research purposes. We categorize our findings into the following categories:

Interactions, which looked into how much the interactive capabilities and settings in the virtual environment could be altered by the researcher. We term our findings from this category into 1. *Programmable*, meaning that researchers can program or otherwise implement their own features without major limitations, 2. *Partially programmable*, meaning there are limited programming or customization options, and 3. *No influence*, meaning there is no way to influence the features.

Customizable Avatars shows the capability of the user to customize their avatars, either by uploading custom avatar models onto the platform, or using the customization options offered within the platform (like choosing different body types, colors, clothing, and accessories).

Data Collection classifies how much freedom a researcher has in collecting data from the platform. Most platforms, such as Rec Room, have limited options or limited tools to help with data collection, and some platforms do not support any kind of data collection. AltspaceVR, though, can be installed on one's own server, making data collection possible.

Saving data collected from participants in the platform is usually not possible. For example, Rec Room presents tools for creating questionnaires, but does not allow saving them. This means that the data must be stored through other means. However, not storing participant data in the VR platforms would not be advisable in any case, as such approaches may not be compliant with privacy regulations (participant data would then effectively be stored by a third-party company). We discuss the implications from this in the subsequent chapters.

External Applications and Sources describes whether any external applications, media, or other files can be integrated into the virtual environment in these platforms. This might be useful for many remote studies, for example, to display study-related information or material.

Finally, **Viability** is our subjective overall assessment of how viable each of the analyzed platforms are for remote VR studies. This is largely based on the analysis of the previous characteristics. This assessment gives an indication of which platforms might be good starting points for considering whether they meet a particular study's needs.

The most popular VR platforms among our survey respondents were VRChat, Rec Room, and AltspaceVR. Based on our analysis of the features of various VR platforms (Table 2), these same platforms stand out from the others as they offer the highest flexibility for researchers. Hence, we provide a brief overview of these three platforms. Other VR platforms that are suitable for research purposes may emerge in the future.

Rec Room is a popular platform with more than 1 million VR users [60]. It allows users to play and create VR games and it is compatible with several operating systems and devices. Users can program simple logic in their custom rooms [9], making Rec Room a powerful platform for studies that only need simple interactive features.

VRChat is a highly customizable platform. Custom worlds can be uploaded into the platform (e.g., worlds built using Unity and the VRChatSDK), offering many features (e.g., spawn points, interactive objects). Prior research used VRChat for studies [51], finding that this was generally a useful approach that, however, came with several limitations. For example, developers with new accounts cannot upload content without earning a certain 'trust' level. Furthermore, there are compatibility issues with some devices and custom content cannot communicate with outside services.

AltspaceVR is a social VR platform supporting several devices (e.g, HTC Vive, Oculus). It allows new world templates and world objects to be created via Unity Uploader, which can be uploaded and shared on the official hub. Due to security reasons, no scripts are allowed. Developers can, however, place so-called "extensions" into their world through external web pages that get translated into 3D content.

---

[9]https://recroom.happyfox.com/kb/article/43-custom-rooms-tools-and-tutorials/, last accessed May 6, 2021

Our analysis of social VR platforms in this section had two goals: first, to gain an overall understanding of the capabilities of such platforms, and whether they are feasible for remote studies in the first place; second, to understand the existing platforms in order to make an informed decision about our case studies and framework, as presented later in the paper. It is likely that these platforms continue to evolve and so their specific qualities will change. Nevertheless, our analysis (Table 2) highlights factors that researchers can consider when evaluating the feasibility of future VR platforms for their studies.

4.2 Distribution of VR Applications

In addition to developing and setting up their VR applications, researchers need to think about how the applications will be distributed to participants and how potential participants can be reached.

4.2.1 Direct Download. Perhaps the most straightforward option is to distribute the application directly to participants, for example, by sending them a download link. Participants then install the application on their own.

It is likely that this approach is relatively effortless and offers the most flexibility, as the application does not need to conform to the rules and regulations of any external services. However, some challenges are likely present. For example, direct installation requires users to install software from untrusted sources (which might limit their willingness to participate) and it might generally be more error-prone than other ways of distribution (e.g., due to lack of support during installation).

4.2.2 Social VR Platforms. As already discussed, social VR platforms might be utilized for remote studies, as study environments can be set up through them. Since the VR application resides within an existing platform, there is no need to explicitly distribute the application. Here, we might utilize the existing user base of the social VR platform, or attempt to attract new users to it.

4.2.3 App Stores. Application stores offer another, potentially effective way to distribute VR applications. Here, applications are published in app stores, such as Google Play[10], Steam[11], or the Oculus Go Store[12]. There are also emerging services that function as a type of app store, like SideQuest[13].

Using app stores to distribute study prototypes poses certain requirements to researchers. Firstly, app stores might require the application to be developed using certain tools and programming languages. Secondly, applications are usually subject to a review and, hence, researchers must fulfill certain criteria regarding data protection, user interface design, etc. Thirdly, publication on app stores are often not free of charge.

At the same time, publishing in app stores is attractive because a large audience can be reached. Indeed, prior research has successfully distributed research prototypes through app stores in more conventional contexts like mobile interaction [26–28]. Furthermore, since the application stores take care of the installation process, it tends to be safe, free of errors, and easy for consumers.

We provide a brief overview of some popular application stores and their viability for publishing applications for research purposes. Our list is not intended to be exhaustive, but rather, to provide a general idea of the possibilities and challenges with app stores. Other app stores might be available or emerge in the future for research purposes.

   Steam is an app store particularly popular with games and VR games that support SteamVR. One caveat of Steam is its strong focus on (commercial) games. There are only a few exceptions,

---

[10]https://support.google.com/googleplay/android-developer/answer/6334282?hl=en, last accessed May 6, 2021
[11]https://partner.steamgames.com/doc/store/creator_homepage, last accessed May 6, 2021
[12]https://developer.oculus.com/distribute/latest/tasks/publish-submit-app-review/, last accessed May 6, 2021
[13]https://sidequestvr.com/, last accessed May 6, 2021

such as software for content creation. Nevertheless, it might be interesting if the research prototype can be embedded into a game, which is particularly popular in the HCI community [15, 37, 61, 63]. A strong advantage of Steam is that it has millions of users and that many HMDs supporting SteamVR exist, hence allowing a large audience to be reached. Researchers need to be aware that publishing takes at least 30 days and has a charge of about 100$ per application.

The Oculus Store is targeted at Oculus devices and, hence, cannot be used by owners of some other popular HMDs. At the same time, the consumer base is fairly large. The store provides clear guidelines for content publishing and the review process takes around two weeks. After review, the content may be published live on the Oculus store or it can be shared via a download key with participants.

The Google Play Store is immensely popular, connecting developers to millions of Android users. The Play Store allows for publishing Android-based VR apps, which can be run on a few HMDs and, more importantly, on any Android smart phone serving as a low-cost HMD in a cardboard. A Google Play developer account is needed, which is available for a one time fee of 25$. One can choose to publish based on several available modes (internal test, closed test, open beta, full publish), called tracks. There are also other options to control the availability of the app, like country-specific availability. These options are very useful for researchers who might want to retain some control over how many people install their app or who they recruit as participants. Overall, the Google Play Store is a highly viable option for researchers, with the downside of not being an option for most high-end VR HMDs.

4.3 Summary

We identified several different ways to conduct remote VR studies. On one end of the spectrum, researchers can produce isolated VR prototypes and distribute them directly to participants. On the other end, researchers can set up their studies in existing VR platforms and make full use of their features. At the same time, there are options in-between, like distributing a custom-built application through app stores, or uploading it to a VR platform that supports custom applications.

These approaches offer a good starting point for our investigation. It is worth noting that there are other considerations involved as well, such as how participants can best be recruited for remote studies. We will get back to these considerations in the later chapters, when we synthesize our findings from the online survey and the two case studies.

With these possibilities in mind, we then conducted the two case studies, which we present next. For these studies, we identified the best ways to run them remotely, which were very different from each other. This also allowed us to report on a wide range of experiences.

5 CASE STUDIES

To complement the online survey as well as our analysis of different approaches, we present two case studies to gather first-hand experience. We selected these two studies as our case studies, because based on their individual needs we concluded that they would require very different approaches. We explain our reasoning in the case-specific sections.

The studies we present are part of independent research projects, in which some of the co-authors of this submission are involved. In this paper, we focus on aspects related to the *study methodology* in the description of the case studies, rather than on the main research questions of the projects and the obtained results. The results are instead discussed in two separate papers [36, 49].
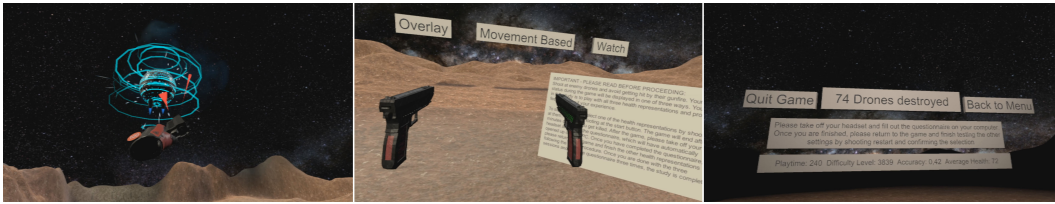
Fig. 3. The VR shooting game and instructions used in Case Study 1. *LEFT:* players shoot drones using their VR contollers as pistols. One of the tested health interfaces, the watch, is visible on the user's wrist. *MIDDLE:* The starting instructions and a menu where the players choose which health interface to use. After their selection, they can begin the game. *RIGHT:* Instructions and menu displayed after playing a session with one of the health interfaces. Study participants repeated this procedure three times.

5.1 Case Study 1: Health Interfaces in VR Shooters

In Case Study 1, we developed a VR shooting game (Figure 3) where study participants attempted to shoot at drones using their VR controllers, and avoid getting hit by the drones' lasers. Our research focus was on investigating different ways to convey the player's *health*. The player's health was reduced each time they got hit by lasers, and would slowly restore if they did not get hit for enough time. The game ended when health was reduced to zero.

We designed and implemented three different interfaces to communicate the player's health status. Study participants played three rounds of the game, each time with a different health interface, and provided feedback.

In this study, the most critical aspects of the remote VR methodology were:

- **Standalone VR application.** We built a custom, standalone VR application. We did not use existing social VR platforms. This choice was motivated by two factors. First, existing social VR platforms did not allow for specific implementations like what we needed. Second, we needed to gather extensive qualitative and quantitative data that was not either possible or convenient if the study was run on an existing platform.
- **Direct download.** We hosted the application on a web server and provided a link to it to participants. Participants then downloaded the application and ran it on their own. We opted for this because it was straightforward and fast to set up.
- **Independent study method.** The study was asynchronous, i.e., participants ran the study on their own any time they wanted. No experimenter was present during the study. We opted for this for two primary reasons. First, the study was about a single-player game where there are typically no observers present. Hence, we hypothesized that playing the game alone was more natural for the participants. Second, we wanted to gather first-hand experience about asynchronous studies, as we anticipate this to be one of the greatest strengths of remote studies, as researchers do not need to allocate time for each participant individually.

5.1.1 Study Description. We investigated whether diegetic representations of player's health status in VR shooting games would improve the sense of presence and add to a sense of danger during intense gaming sessions. As a non-diegetic baseline condition, the player's health was displayed as a health bar, which was visible on the screen at all times. As diegetic conditions, we implemented a wristwatch displaying the player's health (players would need to lift their wrist and really look at the watch) (Figure 3, left) and a movement-based method, where hurt players moved slower and their firearm was shaking, to simulate the condition of being physically hurt. We hypothesized that these diegetic conditions would improve certain aspects of the gaming experience.

Study Design. We built a custom VR shooter game using Unity and SteamVR. We chose these tools because they are by far the most popular ones to use in creating VR applications, and there is a considerable amount of relevant resources and support online. Furthermore, Unity and SteamVR support many of the higher-end consumer HMDs, like the HTC Vive Pro, which was the equipment we had for development. All these points considered, Unity and SteamVR were reasonable choices for building our standalone VR application and for conducting our investigation of this approach.

In the game's main menu, instructions were displayed for the player. They could choose any of the three health interfaces and start the game (Figure 3, middle). In the game, drones appeared around the player which attempted to shoot at the player. The player's task was to avoid getting hit and try to shoot as many drones as possible. The game session was set to last for four minutes. To support players of different skill levels, we implemented a dynamic difficulty system that adjusted various factors based on how well the player performed: the rate at which the drones appeared, how spread out they would be, and how fast and frequently they would shoot at the player. Also, we added a curve to player health – a typical trick in videogames – so that players would not die as easily as the situation might suggest. These were important factors in ensuring that players of all skill levels would experience a challenge and the thrill of being low on health, which in turn was important for our research goals. At the same time, the dynamic difficulty and the curved health would not let players die very easily and frustrate them. Players regained their health during the session if they remained unhurt for a brief period.

Participants played a short session using all three health interfaces. After each session, a feedback questionnaire was automatically opened on their PC on their default browser, and the game instructed participants to take off their HMD, fill in the form, and then return to the game (Figure 3, right). We also used the form as a means to log data from the game: we pre-filled some of the form questions with game data and prevented users from modifying them.

Even though current research suggests filling questionnaires inside VR [3, 54], we opted for an external questionnaire. We did so because our questionnaire also contained qualitative, open-ended questions that were important to our research questions. These questions simply could not have been reasonably answered in VR without a keyboard. Additionally, the questionnaire was long and would have been tedious to fill in VR. The questionnaire after each session included custom statements and open-ended questions as well as the Game Experience Questionnaire [33] and the Slater-Usoh-Steed presence questionnaire (e.g., [55, 62]).

Using external questionnaires outside of VR brought further benefits. First, we could use an existing survey platform. These are easy to set up and are very reliable. Second, we did not have to directly store anything from the VR application, as all data was gathered through the questionnaire. This was easier and safer than logging and transmitting data from the application.

When participants filled in the form for the third and final time, they ticked a checkbox stating that they had finished playing with all three health interfaces. Participants were then presented with a brief set of additional questions. This set included basic background questions, and we also asked them to answer three statements regarding their experience of participating in a remote VR study. Additionally, there was an open text field for free comments. We did this to gain insight into how the participants felt about the remote study and whether they had any issues.

We took special care to provide clear instructions to participants and support them throughout the study process. In our advertisement, we provided a brief overview of what the study would contain. In our recruitment email, we had instructions for downloading and running the VR application as well as the high-level steps they should take in the study. Inside the VR game, we had a text box containing the same instructions, which the users would see immediately after starting the application (Figure 3, middle). Further, there was a brief explanation about each health interface before they started the game. After the game, there were instructions to take off the HMD and fill

in the questionnaire on their computer, and then return to the game (Figure 3, right). The same reminder was added at the end of the questionnaire, that told participants to return to the game.

In addition, we set up a separate issue report form, that prospective participants were encouraged to submit in case they experienced any problems with the game or the study. This issue report form was added in the advertisements as well as in the emails that contained the download link.

Recruitment. We did two rounds of recruitment. In the first round, participation was voluntary, i.e., no rewards were offered. We advertised the study through channels we had identified as suitable for this research, such as Reddit. We expected Reddit to be a good recruitment platform, as a large number of our online survey respondents came from there, and Reddit also contains a large number of subreddits dedicated to VR technologies and HMD owners. We also advertised the study to a practical course about VR technologies and programming that took place at the same time at a local university. Some of the students owned a private HMD, and some had been loaned an HMD by the university. The HMD loans were for several months, and the students installed the setups in their homes as they would install their privately owned HMDs. The students had already been using the HMDs for some time before our study. As such, we concluded that they would reasonably count as HMD owners for our case study.

In the recruitment call, we included instructions for the study procedure (even though all the instructions were in the game as well), because the study was designed to be run fully independently and no experimenters were present at any point. It was therefore important for participants to know exactly what to expect in the study. The call also included a direct download link to the game as well as a link to the issue report form.

In the second round, we advertised the study through the same channels with a 10€/12$ reward, paid via PayPal. We removed the download link from recruitment calls and instead set up a separate registration form where participants provided basic user information as well as an email address. We then sent the study instructions along with the download link to the registered participants. We did this to retain control, so that no one would try to cheat (e.g, by completing the study several times), and that we would not be flooded with too many participants who all needed to be compensated.

This way, we recruited 24 participants (18 male, 5 female, 1 undisclosed). Their average age was 24 (SD = 7.9). The participants used a variety of different VR devices: HTC Vive Pro (7), Valve Index (7), HTC Vive (6), Oculus Quest (2), Oculus Rift S (1), and Asus Windows Mixed Reality Headset (1).

The participants also reported their estimates on how much they play VR games on an average week, and how much they play other digital games. With regards the VR gaming, 12 participants (50%) reported that they play 0–5 hours on an average week. Seven (29%) estimated 5–10 hours, four (17%) estimated 10–15 hours, and one participant (4%) estimated that they play 20+ hours per week. With regards to other digital games, seven participants (29%) estimated that they play 0–5 hours a week, eight (33%) estimated 5–10 hours, three (13%) estimated 10-15 hours, two (8%) estimated 15-20 hours, and four (17%) estimated that they play 20+ hours per week.

5.1.2 Participant Experience. In addition to collecting data to answer our study-specific research questions, we gathered impressions on how the participants felt about the remote study. At the end of the study, participants answered statements about their remote study experience. We asked them to respond to three statements on a 7-point Likert scale, where 1 = strongly disagree, 4 = neither agree nor disagree, and 7 = strongly agree.

Participants were very positive about their experience, strongly agreeing with all three statements (MD = 7). They felt that **(1)** they were comfortable participating from home, **(2)** the instructions for how to run the study independently were clear, and **(3)** they did not need anyone in the same room

to consult. In the free comments section, no issues or negative points regarding the study were brought up. One participant even commended us for a "smooth remote study".

The positive feedback from participants is encouraging. It supports our findings from the online survey, where participants felt comfortable about the thought of remote participation. We were somewhat surprised by this. Rather, we expected that the fully independent study procedure might make some participants stressed or uncomfortable, but this did not seem to be the case. This strengthens our belief that our procedure was well thought out and our instructions prior to and during the study were clear (as also reported by the participants). Still, it is possible that the fully independent procedure drove away some potential participants in the recruitment phase.

In addition to the clarity of instructions, we believe that convenience and familiarity played key roles in the positive experience. In remote studies, users do not have to spend money and time traveling to a lab and they do not have to get familiar with new equipment or meet with new people. In contrast, users can participate from home. This familiar setting likely provides support and safety. They also use their own equipment that they are familiar with before the study.

5.1.3 Experiences and Discussion of Running a Remote Study. In the first recruitment round, we faced a major challenge with getting a sufficient number of participants. Logically, this was because we did not compensate participants, and we required a SteamVR-compatible HMD, so low-end devices were not suitable. A total of 10 people participated in the study. Five of them came from the practical university course, three came from Reddit, and two from other channels.

In the second round, we immediately received 33 registrations overnight, which resulted in us having to close the recruitment form. All who registered were sent the instructions, and 15 of them participated. Therefore, offering a reward solved the recruitment issues. Had we kept the registration form open for longer, we would have likely been able to recruit many more participants, despite the pre-requisite of a SteamVR-compatible HMD. A separate registration form also seemed to be an good choice for controlling the number of participants. Direct access to the study could result in issues for researchers, who often have a limited budget for rewarding study participants.

Since our work in this paper is primarily focused on investigating the different methods and considerations for remote studies, we changed the rewarding scheme mid-study to investigate its effects on recruitment. As expected, offering a reward made recruitment significantly easier. Based on our experimentation, researchers can make a more informed decision about whether or not to offer a reward. If researchers do not offer a reward for their study, then they should prepare to advertise more actively and expect to gain a smaller sample of participants.

One of our initial concerns was that we would suffer from a high drop-out rate (participants quitting mid-study). However, out of the total 25 people who ran our VR application, 20 participants (80%) completed the study in full. Four participants (16%) completed two of the three conditions, and one participant completed only one condition. We see this as a very high completion rate for an entirely remote and independent study procedure that took approximately 30–40 minutes. We believe that this success was due to our efforts to provide as clear instructions as possible for the participants to perform independently, and to make the transitions between study phases as effortless as possible (e.g., moving from the game to a pre-opened, pre-filled questionnaire). The importance of clear communication and instructions was also highlighted by Bramley et al. [10], who conducted a remote study using Google Cardboards and the participants' own smartphones.

However, the number of second-round participants (15, of which 12 completed the study in full) was low compared to the number of registered prospective participants (33). It is difficult to say why so many registrants did not follow up in the end, but our experience would indicate that researchers should prepare for this happening.

Three people reported issues through our issue report form. Those were of technical nature. Two participants, who both had Valve Index controllers, reported that they were not able to shoot using the triggers, but they were able to shoot by pressing other buttons. One participant reported that their pistols were slightly misaligned with their hands. The reported issues were minor, as participants were still able to complete the study. Still, such instances are testament to the challenges with remote studies, as things that researchers typically have control over (e.g., calibration, testing, solving issues on the spot) are outside of their grasp.

Our choice to use external questionnaires outside of HMD use made collecting data very easy. This was largely motivated by our need to collect open-ended data which was not convenient while the user is in a virtual environment. Typically, though, it is better to have questionnaires filled in VR [54], which in turn presents considerations for how the data is stored, and what limitations exist. We return to this point in section 6.2, where we discuss different possibilities and workarounds for collecting data, and how this might inform the design of a remote study.

5.1.4 Conclusion. Using our approach, and largely thanks to carefully designed instructions, we completed the study as planned. Aside from the initial troubles with recruitment (which we overcame later), we did not experience any issues. Additionally, the participants gave very positive feedback about their remote study participation.

As we could expect based on our survey data, our participants were dominantly male (75%). Although there is a prevalent gender imbalance in HCI studies in general [13], this bias is still noticeably strong. The average age of our participants was 24 (SD = 7.9). This is indeed not high, but it is likely no lower than what HCI studies typically have. While we are not aware of an average participant age across HCI studies, we do know that study participants are typically students [13], and they likely fall in this range. It is also worth noting that our standard deviation was relatively high (7.9), with some participants being over 40 years old. With respect to the participants' gaming habits, most were rather moderate with their VR use (0–5 hours or 5–10 hours per week). Instead, they were more actively gaming on other devices (e.g., PC, consoles). This is not surprising, considering the potential fatigue in VR gaming as well as the fact that VR gaming is not yet as mature as other forms of digital gaming. In any case, our participants seemed to represent a relatively balanced group of moderate games and hardcore gamers (judging from their play time).

5.2 Case Study 2: Investigating Proxemics in VR using Rec Room

In Case Study 2, we set up a VR environment (Figure 4) where two participants completed collaborative tasks. Our research focus was on investigating how the gender of virtual avatars affects the interpersonal distance (IPD) between people. We measured the distance between the participants at specific points during the study.

We designed a virtual environment in Rec Room and implemented two different collaborative tasks, as well as implemented a sophisticated system to track the distance between participants. Participants switched their avatar genders halfway through the study.

In this study, the most critical aspects of the remote VR methodology were:

- **VR application built within an existing VR platform.** We built a custom VR environment in Rec Room, using the tools provided by the platform. Because our study was about proxemics and not about specific interaction techniques or user interfaces, the social VR platform Rec Room had sufficient functionality for our study and it was easy to set up. A significant benefit was that because our study was about social interactions between two people during collaborative tasks, using a social VR platform made sense because multi-user connectivity and the required interactive functions were already there.
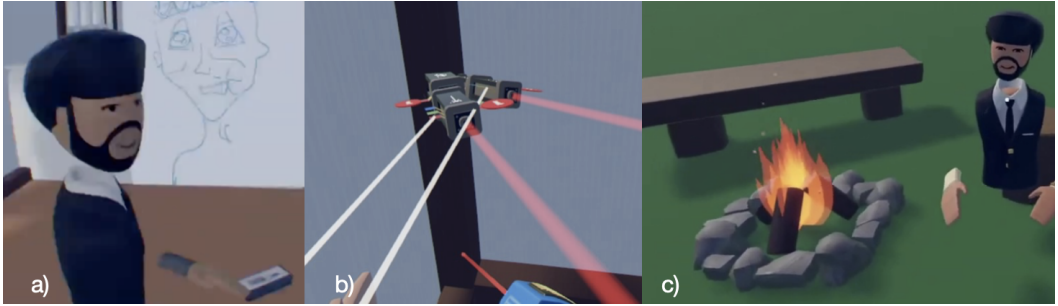
Fig. 4. Proxemics study (Case Study 2). a: Players playing inside the VR game. *b:* Range finder feature of Rec Room (used to calculate the distance between players). *c:* Conducting interviews at the end of the study.

- **Distribution through the VR platform.** Because our application was built inside Rec Room, there was no need to download or install an application. Instead, our study was accessible through Rec Room. This also gave us access to Rec Room users as potential study participants.
- **Remotely guided study method.** A remote experimenter was present during the study to recruit users and to set up and manage the study environment in each session. We needed an experimenter present to record some of the data that we collected, as Rec Room did not allow storing the data directly. We furthermore concluded that managing collaborative tasks between two participants would have been very difficult without having having an experimenter manage the situation in real-time.

5.2.1 Study Description. We investigated proxemics in VR using Rec Room. The proxemics theory studies the interpersonal distance between humans during interaction [17]. We explored to what extent this theory applies to VR. In particular, we investigated whether the relationship between people (strangers and friends) as well as gender affects proxemics.

Study Design. Our earlier analysis of social VR platforms suggested that the most viable options for a remote study were Rec Room, AltspaceVR and VRChat. These platforms were also the most popular among the respondents of our online survey. From these, we chose Rec Room because our testing showed that we could implement all the necessary features with it, and the platform's design tools were easy to use.

We designed a collaborative two-player game in which participants had to solve puzzles together to unlock rewards (Figure 4a). To understand the influence of the relationship between people on proxemics, we recruited both pairs of friends as well as pairs of strangers. To additionally investigate the role of gender, the study consisted of two phases. In the first phase, participants collaborated on the puzzle-solving task with an avatar matching their gender. In the second phase, participants were represented by an avatar of the opposite gender. The study consisted of two sessions. After the first session participants were required to go back to their own room and change their avatar gender and then return to join for the second session of the study.

The experimenter was present in the room throughout the study. The experimenter embodied their own avatar in the study space, provided instructions to participants and oversaw the procedure, very much like an experimenter would do during a lab study.

The distance between the players was logged using the Rec Room feature "Range Finder" (Figure 4b), that can track the distance from a certain point to a player. However, because Rec Room does not allow storing such information, we had to implement a workaround. We set up the Range

Finder data on a virtual display in Rec Room and recorded the data using screen capture. We then used Optical Character Recognition (OCR) to extract the data.

After the players finished the game, they filled in a demographics questionnaire. As stated in our analysis of social VR platforms, Rec Room did not allow storing questionnaire data, nor would it have been advisable due to privacy protection regulations. Instead, the questionnaire results were displayed in the virtual environment, from which the experimenter noted the answers down manually. Finally, the experimenter interviewed the participants inside Rec Room (Figure 4c). Different rooms were designed for each purpose so participants played the game in one room and completed the interview in a separate room. Participants were compensated with 5€ via PayPal.

Recruitment. We distributed a call for participation via various social media platforms, university mailing lists, in the Rec Room community on Reddit, and via Facebook. In addition, we reached out to people active in the Rec Room community. In particular, one moderator from the Reddit Rec Room group advertised the study via social media (Facebook). One Rec Room tutor, who hosts a tutorial class on circuit and logic design inside the platform, suggested this study to his followers. Furthermore, we asked participants to advertise the study to friends who might be interested. This led to additional participants. Finally, we recruited active players in Rec Room on the fly. This turned out to be an effective method, since those players were already in Rec Room and were often available for participation.

There has long been gender bias in recruiting participants for VR studies [48]. We also struggled to find female participants for our study. This led us to have an imbalanced gender ratio and, therefore, we could not conduct studies between three gender groups (male-male, male-female, female-female) as planned initially. Eventually we had only male-male pairs (15) and male-female pairs (5). Therefore, we had 40 participants (35 males) with an average age of 23.7 years (SD = 7.75). Aside from the gender imbalance, we could successfully recruit 40 participants, and it was particularly useful to recruit participants directly from Rec Room. One drawback was a difference in time zones between the experimenter and the participants, and thus the experimenter had to conduct some sessions at unusual hours (e.g., very late at night).

5.2.2 Participant Experience. In addition to collecting data to answer our study-specific research questions, we gathered impressions on how the participants felt about the remote study. Hence, in the interview at the end of the study, we asked them questions about their experience.

All 40 participants were overall very positive about the experience. All of them stated that they felt comfortable about participating in a remote study. Furthermore, 39 participants stated that the remote study in Rec Room was a novel experience to them; one participant mentioned to have previously attended a focus group in Rec Room. We also asked them if they faced any challenges with participating or understanding instructions, but no such issues were reported. We also assessed the study sessions using the video recordings, and we did not identify any moments where participants would have been distracted by external factors, such as real-world noise or interruptions by other people.

Similar to case study 1, it is likely that participants felt positive about the experience due to the convenience (participate from home, no need for travel or other inconveniences) and familiarity (familiar and safe home setting, familiar equipment and platform, prior VR experience with the same setup). The difference to the first study was that we used Rec Room with a remote experimenter, which also seemed to function very well from the participants' perspective. It is notable that Rec Room is a social platform where participants likely had many interactions with other people before. Therefore, interacting with a stranger (the remote experimenter as well as the other participants in the stranger condition) was not new to them.

5.2.3 Experiences and Discussion of Running a Remote Study. In total we recruited 40 participants in pairs. As many friends play together in Rec Room it was easy to recruit friends in pairs. It was also easy to recruit strangers in pairs. For recruitment, the experimenter would generally join playing rooms and approach other players about the study. Since each game room consists of many players, both friends and also strangers playing altogether, recruiting was easily possible.

Our main challenge was the balanced recruitment of male and female participants. Ultimately, we failed to do this, as recruiting female-female pairs proved to be difficult. Even though we otherwise ran a very successful remote study, the lack of female participants limited our findings regarding gender differences. While we also faced a bias towards males in the first case study, the issue was more clear in this second study. This is likely due to two factors. First, we needed participants to attend the study in pairs; finding two female participants for one session as opposed to just one female participant is naturally more difficult. Second, according to our survey the general pool of HMD owners (and likely Rec Room users) is male-dominated.

One important aspect to consider is that the Rec Room community consists of a large number of users below the age of 18. Hence, proper verification of the participants' age should be put in place upon recruitment. At least, participants should be told that they are required to be at least 18 years of age to be eligible to participate. But this also makes it easy for researchers to recruit underage participants if required for the study. In this case, researchers would need to find ways of obtaining consent from participants' parents.

One challenge during recruitment and conducting the study sessions was that potential participants were spread across the globe. Therefore, some participants were not available at regular working hours for the experimenter. As a result, the experimenter ran some sessions in the middle to accommodate for different time zones.

We planned for position tracking in VR to measure the interpersonal distance of participants. Rec Room allows the position of each player to be visualized; however, logging this data was not possible. We worked around this by recording the visualization during the study and then using optical character recognition to extract the numbers. This worked well in the end, but required some experimenting to get right, so that OCR would produce the correct output. Depending on the study being conducted, implementation challenges are likely to occur when using existing VR platforms, as their features can be limited.

We similarly had to work around data collection limitations and regulations with our demographic data, which we collected via questionnaires. The experimenter noted the data down manually. This was not an issue for a short set of questions that we used, but might be riskier and more time-consuming if a larger amount of data was gathered in this way.

Another design challenge specific to Rec Room was the limited ink[14] available for designers. The pen system in Rec Room allows creators to design a room and this has an ink limit to facilitate robust processing, i.e., the rooms have a limit to how many objects they can contain. Thus, when designing the room, we had to plan economically.

The study was interrupted a few times when the participant's game crashed and they left the room. Additionally, since the experimenter was also virtually present wearing an HMD device, it was difficult for the experimenter to check their personal notes during the study.

5.2.4 Conclusion. Thanks to some creative workarounds, our study procedure was completed as planned. We had no problems during the study sessions, the participants' were positive about their experience, and our data collection methods worked as intended. Recruiting participants turned out to be easy in several ways. Most importantly, the option to directly recruit participants from inside Rec Room was very helpful.

---

[14]https://rec-room.fandom.com/wiki/Maker_Pen

As noted earlier, though, it was critical for our research to also recruit female pairs as participants, and this was a major challenge. This showcases the limitations in the demographic of HMD users, although in most studies the issue should not be as pronounced as here. Two minor inconveniences were that some participants were in far-away time zones, and that Rec Room had limited data collection features, so we could not record the exact distance between participants directly. Researchers should be prepared for creative workarounds in such cases.

5.3 Discussion on the Validity of Remote VR Studies

There are some challenges unique to remote studies that might put the validity of their results into question. In particular, there are two such factors: lack of control (e.g., possible external distractions and misunderstandings), and device heterogeneity (e.g., unforeseen technical issues, different experience as a result of using different technology) [25, 27].

We believe that the case-specific results acquired using our remote approaches are reliable. In case study 1, the questionnaires that participants filled in three times, included a timestamp. These timestamps reveal that every participant filled in all three questionnaires in the same session, with a consistent 10–15 minutes between them. We also did not receive duplicate data from any condition, nor did participants report any procedure-related issues. These findings suggest that participants successfully completed the study once they began, and did not suffer from major disruptions. In case study 2, neither our recordings nor observations during the study suggested any external disruptions among the participants. The participants followed instructions and answered questions without trouble. In addition, existing research has studied the validity of remote VR studies and found the results comparable to lab studies [46, 50].

Even though our study procedures went smoothly, remote studies might still suffer from disruptions more than traditional studies, as experimenters are not present at all or at most they monitor the situation virtually. Issues might arise especially if a complex study design is needed, if the study is long, or if the participants need to be particularly alert. Additional mechanisms could be used to assess the validity of the study sessions. This could be done, for example, by recording the participant's screen, gaze, or movement, which could be analyzed post-hoc to evaluate the participant's behavior and attention. This, in turn, has ethical implications: participants must consent to the logging of behavioral data, and sufficient safety measures must be carried out as real-time logging of data from the user's machine might present security and privacy risks.

Another reliability-related discussion is whether device heterogeneity might play a role in the results [25, 27]. Remote studies in other areas have also suffered from technical issues stemming from, e.g., different Android versions on smartphones [27]. The extent of this is likely specific to the needs of the study. In our case study 1, participants could only use SteamVR-compatible HMDs which, even though there are many such devices, did somewhat unify the technologies used (e.g., in terms of tracking space and controllers). A few participants reported minor technical issues, but all could still complete the study. This was partly because our study was also not particularly dependent on the controllers, as we only needed them for aiming and firing (one button). Similarly, our case study 2 was focused on proxemics and movement rather than specific interactions. Here, we also believe that the exact HMDs used were not as critical.

Studies relying on more complicated controller input might need to pay more attention to the different devices. Similarly, the resolution of HMDs can vary, which affects how well users can perceive things in the virtual environment. This is critical to many studies, so again the effect of the HMD should be accounted for. In any case, researchers should consider the capabilities and functions of available HMDs, and evaluate their possible effects on the study. This could be done beforehand to exclude certain devices from the study, or post-hoc to test for potential effects.
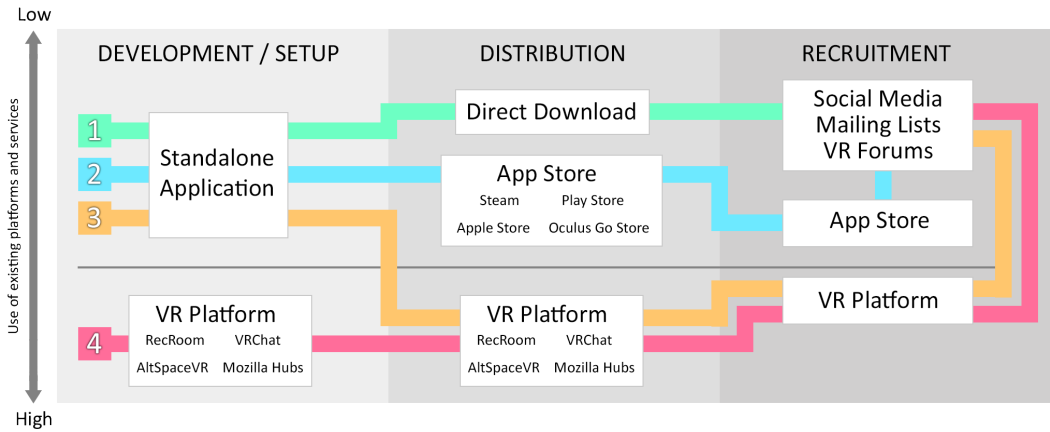
Fig. 5. We envision four high-level procedures to conducting remote VR studies. The organization is based on how much the approach relies on existing platforms and services, Path 1 being the most independent in this sense, while Path 4 is fully built on an existing platform. Existing platforms and services bring benefits such as existing design tools and assets; ease of use; speed of setting up virtual environments; and recruiting from an existing user base. At the same time, they are (sometimes significantly) more limited than approaches that are more independent in terms of, e.g., interactive and data collection capabilities. *PATH 1:* Standalone VR application that participants download and install independently. *PATH 2:* Standalone VR application that is published in an app store. *PATH 3:* VR application that is uploaded to an existing VR platform, which can also be used to recruit participants. *PATH 4:* VR application that is built using the built-in tools of an existing VR platform, which can also be used to recruit participants.

6 A FRAMEWORK FOR RUNNING REMOTE VR STUDIES

In this section, we synthesize the findings from our online survey, the analysis of different approaches (standalone, app store, VR platform), and two case studies into a framework for remote VR studies (Figure 5). We discuss the strengths and weaknesses of each approach, and when each approach might be appropriate. The framework is meant to assist researchers in planning their own remote VR studies.

6.1 Four Primary Approaches to Remote VR Studies

We distinguish four primary approaches to conducting remote VR studies, which we refer to as Paths 1–4. The most noticeable factor in the framework is the extent to which each path utilizes — and relies on — existing platforms and services. Path 1 is the most independent, with Paths 2 and 3 being somewhat reliant on existing platforms, and Path 4 being fully conducted on an existing platform. Generally, using existing platforms and services makes some parts of the study easier and offers some unique benefits, but may also introduce significant limitations. In the following, we present these four approaches and discuss their differences, strengths and weaknesses.

6.1.1 Path 1: Standalone Application + Direct Download. In Path 1, which we experimented with in case study 1, researchers build a standalone, independent VR application, which participants download and install themselves. The study can then be advertised through various channels. In our experience, VR-related forums (like VR subreddits in Reddit) worked best in attracting HMD owners.

The strengths of this approach are that independent applications can have the most extensive functionality and the most options for collecting data, because the applications do not need to

conform to the limitations and regulations of app stores and VR platforms. In our study, we logged extensive quantitative and qualitative data with relative ease.

Still, this approach might not be ideal for every development platform. For example, installing external applications for Android devices requires extra steps, such as turning on developer mode, which not every consumer is willing or even capable of doing. In our case study, this worked well, as participants only needed to download our game package, after which they could immediately run the game through an executable.

Researchers should consider whether or not they should distribute the download link openly. In case participants are rewarded, researchers might consider mechanisms to control that not too many participants take the study (e.g., so that their budget is not exceeded), and to reduce the chances of people "hacking" the system (e.g., by completing the study and claiming the reward multiple times). Researchers could, for example, set up a website for pre-registration, and only registered participants would receive the download link. In our case study, this approach worked well, although it is likely that not all pre-registrants will complete the study.

We believe Path 1 to be particularly suitable for independent studies, where participants go through the study procedure alone, without consulting an experimenter. In our case study, this approach worked extremely well, although the procedure should be planned meticulously. Extra care should be paid to providing clear instructions. In best case scenarios using Path 1, the VR application is distributed and advertised, after which the entire study runs on its own.

It is certainly possible to run experiments with a remote experimenter through Path 1, but this requires additional steps like agreeing on specific time slots with each participant. Depending on the study, a remote connection to the VR application would need to be developed for the experimenter, or a connection would need to be established via other means like video conferencing tools (which would, in turn, limit how much control the experimenter has and what insights they can gather). However, there is clearly demand for multi-user infrastructures which are being developed [21], and at some point we expect such infrastructures to become available that are accessible and easy to set up for remote study purposes as well.

6.1.2   Path 2: Standalone Application + Publishing through App Store. In Path 2, researchers build a VR application and publish it in an app store. In some cases, it is possible that the app store itself attracts participants, but researchers should be prepared to reach out to potential participants through additional channels.

The main strengths of this approach are that distributing and installing the application is easier and less error-prone (as opposed to independent installation), and might help in reaching a wide and diverse audience. App stores have been commonly used in other research areas, for example, in mobile interaction [26–28], where large participant numbers were reached. Opportunities for data collection are still relatively extensive, although somewhat more limited than in Path 1, as the application must conform to the rules and regulations of the app store.

Based on our analysis of existing app stores as well as own experience with publishing a VR app on Google's Play Store, there are considerable differences between app stores in terms of how easy it is to publish apps (some require a review by the provider) as well as how much effort and time they require (for some of them it takes up to 1 month to receive feedback). Researchers should inform themselves prior to choosing a particular app store.

For the same reasons as Path 1, we believe that Path 2 is more suitable for independent studies; in the best case scenario, studies using Path 2 could also run on their own after being advertised.

6.1.3   Path 3: Standalone Application + Upload to VR Platform. In Path 3, researchers build a VR application and upload it to an existing VR platform, such as VRChat. The study can be advertised through the platform itself, but also through conventional channels.

The main strengths of this approach come from the social aspects of the VR platforms. They are built for multiple users to connect to, and interact with, each other. As such, many things critical to some user studies are already there or can be implemented with little effort. For example, experimenters and additional users can connect to and interact with the environment without additional effort. In addition, the VR platforms serve as potential recruitment channels. Experimenters can look for participants within the platform and recruit them on the spot. In this way, we were able to recruit participants directly from Rec Room in case study 2. Similarly, Saffo et al. [51] recruited participants from VRChat in their remote VR study. These strengths make Path 3 well-suited for remotely guided studies where an experimenter is present, and perhaps particularly suitable for studies where multiple participants are needed simultaneously.

Publishing custom applications in VR platforms requires the application to follow certain regulations, and requires that the application is implemented correctly (e.g., that it uses specific APIs and libraries). This somewhat limits the features that can be included in the application and what kinds of data can be collected. On the other hand, connecting an experimenter allows them to easily conduct in-depth observations and interviews, which are more difficult with Paths 1 and 2. Using a VR platform also helps in overcoming compatibility issues.

6.1.4 Path 4: Setup directly in VR Platform. In Path 4, researchers build the VR "application" directly within an existing VR platform, such as Rec Room, using the platform's built-in tools (like we did in case study 2). The study can be advertised through the platform itself, but also through conventional channels.

Similar to Path 3, the strengths of this approach are that experimenters can easily connect to the study sessions and the platforms serve as excellent recruitment channels. A unique strength of Path 4 is that the tools offered by the platforms make setting up study environments fast and easy, as opposed to implementing custom VR apps. For example, Rec Room offers a simple user interface for building rooms inside the platform, adding and modifying objects, and setting up interactions. Using such tools requires little to no technical expertise, making this path particularly attractive to researchers from non-technical fields.

Still, the possibilities offered by VR platforms are limited, as they are rarely built for research purposes, but for simple social interactions between users. Therefore, when considering which path to take, researchers should look into existing VR platforms to see whether any of them offer the required functionality. A related limitation is that VR platforms offer limited ways to collect (especially quantitative) data from users. Still, it is often possible to build questionnaires within the platforms (like we did in Rec Room). Moreover, like in Path 3, connecting an experimenter allows in-depth observations and interviews to be conducted.

6.2 Choosing the Best Approach for a Remote Study

The four approaches balance different needs for user studies. While most approaches can be adapted to fit different studies one way or another, there are still general considerations that can be used to identify the best approach. It is also worth noting that for certain VR studies, remote approaches might not be feasible at all, or they might present significant challenges.

As discussed in the previous subsection, the extent to which the different approaches rely on existing platforms and services has many implications, for example, on what interactive capabilities and virtual environments can be implemented in the study, what kind of data can be collected, and how much effort the implementation is and how much technical expertise it requires. Below, we discuss these perspectives for evaluating which remote approach might be the most feasible for a planned study:

6.2.1 Consider Your Requirements for Interaction and the Virtual Environment. In deciding which approach to take with a remote VR study, researchers can consider the interactive features that their study requires, and investigate whether existing VR platforms meet their criteria. Similar considerations should be applied to the virtual environment in which the study should take place.

We believe that Path 4 is generally the easiest path to take, and, therefore, it might be useful to evaluate whether Path 4 is a feasible option in the first place. Path 1 offers the most control over what interactive features can be included, with Paths 2 and 3 situating in the middle, being also dependent on the targeted app store or VR platform.

6.2.2 Consider Your Requirements for Data Collection. Remote studies in general are more limited than lab studies, and collecting data through specialized hardware such as biosensors, electroencephalography (EEG) and electromyography (EMG) is very challenging. Despite current technology making a greater range of data available (such as eye tracking data), additional hardware requirements constrain remote studies. Hence, remote studies are not always feasible.

Setting specialized hardware requirements aside, more typical data can certainly be collected remotely. Much like with with interactive features, Path 1 offers the most control over what data can be logged from the application. Path 2 also offers much control over logged data, although some app stores might impose some limitations. Paths 1 and 2 are therefore well suited particularly for gathering quantitative data.

Because VR platforms limit data logging – for privacy and safety reasons – Paths 3 and 4 are less ideal for collecting quantitative data. However, some limited options may still be available, and also some workarounds exist. For example, as demonstrated in case study 2, we were unable to log participant positions from Rec Room. We obtained the required data using screen capture and OCR to extract the data. This was a successful workaround for us, yet required additional effort from us in terms and planning and experimentation, and we also had to purchase additional OCR software.

In contrast, with qualitative data the advantages may be flipped. Because Paths 3 and 4 are well suited for studies where an experimenter is present, they are — by extension — suitable for collecting qualitative data through observations and interviews. Collecting qualitative data through Paths 1 and 2 is certainly possible, but may require additional effort. For example, for standalone applications, the possibility for a remote experimenter to connect to the same session as the participant must be implemented, or they must use additional software to connect to the sessions. Video conferencing tools may be helpful here. However, despite their ease of use they still add additional steps to the procedure and might not be enough to capture all details in the session.

Therefore, studies relying heavily on quantitative data (or the collection of large amounts of data) might benefit more from Paths 1 and 2. For studies with a qualitative focus, Paths 3 and 4 could be considered. We again emphasize that these are not set in stone. For example, with Paths 1 and 2, video conferencing tools, screen sharing tools, etc. could be used to connect with participants, but they are still limited and require extra steps in the procedure.

6.2.3 Consider the Advantages and Drawbacks of Independent and Remotely Guided Studies. As already discussed, independent studies likely support quantitative data, while remotely guided studies might support qualitative data. There are also other considerations to these two approaches.

For independent (asynchronous) studies, the benefits are that since the actual procedure requires no resources, the study can run day and night and potentially reach a large number of participants. For certain studies, it is also considerably easier to develop a VR application if one does not have to worry about functionality surrounding the experimenter. However, due to the lack of an experimenter, independent studies might be more prone to malfunctions (e.g., bad calibration, hardware issues) and misunderstandings (e.g., not following the procedure correctly), or participants might simply be overwhelmed or discouraged by the procedure.

In contrast, through a remote connection, experimenters can instruct participants and oversee that the procedure is followed accordingly. Especially, they can tackle any unforeseen events and provide support if problems occur, and conduct observations and interviews. Remote guidance might be particularly desirable if the participants need to learn or train new things, if the procedure has several phases with instructions, or if the study has multiple things running in parallel, like several participants. A case in point is our case study 2, where participants attended the sessions in pairs. This study would have been very difficult to run without an experimenter who was required to monitor each session and instruct participants accordingly, especially when considering the limitations of the used VR platform. The obvious drawback is that an experimenter is then tied to running each study session, which takes time. Experimenters might also face challenges with different time zones; flexibility with schedules might be needed to reach participants located in other countries.

6.2.4 Reflections on the Longevity of the Framework. Here, we reflect on how we believe our framework, the different approaches to remote VR studies, and our considerations might evolve in the long term. It is inevitable that the VR landscape, the tools, and the user base will evolve. However, we have formulated most of our findings and considerations in such a way that they will still hold despite the changes that we expect to see in the future.

We have only recently arrived at a situation where a notable enough number of people own head-mounted displays. But VR technologies and the related tools are still relatively immature, and the number of people who own HMDs is still marginal to that of personal devices, such as smart phones and smart watches. We believe that in the future, remote VR studies will become a more feasible method for a wider range of studies, as VR technologies develop and a larger audience becomes equipped and acquainted with head-mounted displays. We learned through this work that HMD owners are still rather homogeneous in certain aspects (e.g., gender, gaming expertise, VR expertise). As HMDs inevitably become a more standard household item, we expect to get more diversity among people who have access to HMDs at home.

Even though there are already a wide range of HMDs available, we expect to see an even larger number of different devices in the future. This can be both a challenge and an opportunity for remote studies. It is likely that some future HMDs will have fundamentally different functionalities and capabilities (e.g., new types of hand-held controllers) from what we might be used to. Therefore, remote VR studies must either restrict the types of applicable HMDs or consider ways to make their study functional with different devices. At the same time, all kinds of studies — lab-based studies included — will have account for different types of devices, which can be an issue especially if researchers do not have access to a large selection of HMDs. Remote studies with participants using a diverse set of HMDs is an opportunity to tackle that challenge.

Through the future advancement of VR technologies, we may eventually have more data collection opportunities in remote studies. However, we believe that studies using app stores and especially studies using social VR platforms will continue to be limited in their data collection. Such limitations are in place for good reasons—to protect the privacy and safety of the users—so this is unlikely to change. Hence, Path 1 (standalone application + direct download) might benefit the most from advanced data collection, which we believe is already the most flexible in this regard.

Data collection limitations aside, it is reasonable to assume that social VR platforms will continue to develop alongside HMDs, and offer more advanced and comprehensive interactive options as well as tools for content creators. Therefore, we believe that Paths 3 and 4 (that utilize social VR platforms) will become a feasible option for a wider range of studies.

7 BEST PRACTICES & LESSONS LEARNED

In this section, we draw upon our results and experiences, and provide some best practices and lessons learned that we hope help researchers while planning their remote VR studies.

7.1 Understand HMD Users and Demographic Limitations

Even though it is possible to reach a large number of participants remotely compared to a lab study, remote studies may suffer from certain demographic limitations. As we learned from our survey, HMD owners are biased towards young males. This is also evidenced by our two case studies, where the majority of participants were male (75% and 87.5%, respectively).

The bias towards male participants was particularly a challenge in our second case study, where we initially attempted to investigate gender differences. However, due to the difficulties of finding female pairs, we had to exclude them and opt for an investigation between male-male pairs and mixed gender pairs. Hence, we received only limited insight into gender differences.

It is worth noting that the bias towards males is not necessarily solely due to the imbalance among HMD owners, but also due to the gender imbalance in suitable recruitment platforms (e.g., Reddit, Rec Room). In any case, researchers looking to utilize HMD owners as study participants should be aware of this limitation, particularly if gender balance is important with respect to the insights of the planned study.

As an advantage, however, online recruitment platforms allow pre-selecting participants based on demographics (e.g selecting participants who own a VR HMD). Recruitment in online VR platforms is easy as experimenters are likely to find available players online.

7.2 Home VR Setups May be Small

According to our survey, many home VR setups are smaller than what researchers might be used to in lab conditions. 29% of respondents reported a small space of up to only 5 m^2, while another 31% reported a space between 5 and 10 m^2. Therefore, it might be challenging to conduct studies that require a large tracking space. This might be especially true for user studies where active and fast-paced movement is expected, where a small tracking space might not only hinder the participant's ability to engage with the experience, but might also put them at risk of colliding with objects in the physical world. At the very least, researchers should be clear about their space requirements and implement mechanisms to ensure the safety of remote participants.

7.3 Recruit Participants via VR Platforms

In case study 2, where we set up our study in Rec Room, an effective way to recruit participants was to look for them from within Rec Room. Since Rec Room users are already using their HMD and are already within the platform, many of them were available for participation on the spot. This indicates that studies utilizing VR platforms (Paths 3 and 4), researchers might want to consider recruiting participants from within the platform.

7.4 Identify the Rules of Subgroups in Forums and Social Media

VR-related groups on various forums and social media platforms, especially Reddit, were also ideal locations for recruiting participants. However, many such platforms not only have their general community rules, but many subgroups have their own rules. Researchers should identify groups where posting study advertisements is allowed, and contact moderators in cases where the rules do not provide the necessary information.

7.5 Set up a Channel for Participants to Report Issues in Independent Studies

In case study 1, we set up a separate issue report form that we linked to in the original advertisement. Although we luckily did not run into any major issues, some participants still used it to report minor things that they had trouble with. Because of the lack of control and knowledge over how participants run the study, we recommend that researchers set up a similar channel for independent studies. This is particularly valuable with issues that altogether prevent a participant from completing the study, potentially making researchers aware of major issues early on.

7.6 Ensure a Flexible Schedule in Remotely Guided Studies

Potential study participants can be located anywhere in the world and hence they might be in a totally different time zone. In our case study 2, where we used a remote experimenter, this meant that some studies had to be run during the night to suit the participants' schedules. Of course, recruiting participants from several continents is not always necessary, but being able to recruit participants from more diverse backgrounds is one of the strengths of remote studies.

7.7 Make Notes Accessible for Remote Experimenters

In our case study 2, an unforeseen challenge was that it was occasionally difficult for the experimenter to check their notes while wearing an HMD. There are several solutions to this. One practical consideration is for experimenters to have the notes *on their person* instead of, e.g., on a table. In the latter case, experimenters may be disoriented when taking off the HMD, or they may be far away from the table, and may, therefore, need some time to locate their notes. A second consideration is that many VR platforms, such as Rec Room, can be used on desktop machines as well. This might be a more ideal way in some studies for the experimenters to connect from, instead of using an HMD. A third consideration is that it might be possible to put the experimenter's notes within the virtual world, so that the notes are only visible to them.

7.8 Consider the Implications of Remote Research and HMD Use on Data Collection

Remote VR studies introduce unique considerations for data collection and storage.

First, participants cannot easily type while wearing an HMD. This can be an issue especially for independent studies where the researcher is not present and where open-ended questions are valuable for feedback. In case study 1, we solved this by automatically opening questionnaires on the user's default browser and having participants take of their HMD. If this is not feasible for a planned study, researchers might consider excluding open-ended questions, or establish an audio connection with the participant to collect feedback orally.

Second, because participants attend remotely, data storage must be handled safely. There are also possible limitations in this regard if using social VR platforms, but there are ways to get around these challenges. In case study 1, we had all data transmitted through external questionnaires; no data was logged nor transmitted from the VR application. In case study 2, participants filled in questionnaires inside Rec Room; however, this data was not (and could not) be stored in Rec Room. Rather, we had the data displayed on the walls of the VR environment, and the experimenter recorded the data manually. In addition, participants were interviewed, which was recorded from the experimenter's machine. No data nor files were transmitted from the participant's machine.

Besides these unique considerations, data collection is always a critical aspect to consider, no matter what data is collected and how [25]. While our work in this paper focuses on understanding remote study procedures and their design, researcher should additionally be aware of accepted practices related to data collection and storage [25], and also be aware of privacy regulations such as GDPR and CCPA [2, 24].

7.9   Review Ethical Concerns and Ensure Transparency about Potential Risks

Researchers are responsible for ensuring that their research follows ethical conduct [16, 34, 35]. In addition to data collection-related considerations, as discussed previously, there are a number of other ethical considerations, some of which warrant special attention when using VR technologies and when using remote study processes.

VR technologies are associated with health and privacy risks [7, 14, 29, 57], for example, through cybersickness [14, 19], physical injuries [45]), and effects on mental health [1, 30]. The privacy risks of VR technology should also not be neglected, as large amounts of sensitive information may be involved, while access to and transmission of this data is often not transparently regulated [7, 20]. These risks imply a need for reflections on the ethical consequences of VR studies, since they can affect participants. Remote VR studies could even aggravate the moral concerns as they are less controlled.

Ethical concerns must be addressed to ensure safety and security of the participants. For example, the absence of an on-site researcher means no assistance to the participants in case of health issues. In case study 1, we had no on-site researcher present. Thus, we had to exercise specific care during the development phase to minimize any risk. For remote studies, having a remote experimenter present during the study can improve the safety of the participants as the remote experimenter can monitor the participants and stop the study if required, similar to how we did in case study 2. Using internet to download study links or participating online may pose security threats to participants. We strongly suggest that researchers make their best effort to minimize such threats by using verified and trusted sources, such as official institution websites, to host study advertisements, information, and materials (like download links), instead of public sources such as links on Google Drive. Researchers can also focus on distributing studies in forums where safety measures exist (e.g., new accounts cannot make new posts, a certain trust level must be earned, or the community can help by marking safe and valid posts with badges or tags).

Therefore, we recommend a thoughtful consideration of all health and privacy risks and their mitigation. Moreover, transparent communication of risks to the participants is necessary to allow for an informed consent.

8   CONCLUSION

We investigated different ways to conduct VR studies remotely, using the participants' own VR equipment. We first conducted an online survey (N=227) to understand HMD owners demographics, their VR setups and usage, and their willingness to participate in remote VR studies. Second, we analyzed existing app stores and social VR platforms to assess how suitable they would be to be utilized in remote VR studies, and what advantages and limitations they might bring. Third, we conducted two case studies to gather experiences from different approaches. Based on these contributions, we derived a framework, guiding researchers planning to conduct remote VR studies.

Our framework features four plausible paths to conduct remote VR studies. We learned that remote studies come with unique opportunities, challenges and considerations. The most defining factors in designing a suitable VR study for each purpose are whether to utilize existing VR platforms or app stores, or to implement a standalone VR application, and whether the study is run independently by participants, or guided by a remote experimenter.

We also identified best practices and lessons learned. Together with them and the framework, we believe that our work helps researchers to identify the best possible approach for their remote VR studies, and to anticipate potential pitfalls. As VR technologies keep advancing and more VR devices find their way into people's homes, we believe our work to be valuable to a growing number of researchers and practitioners.

REFERENCES

[1] Frederick Aardema, Kieron O'Connor, Sophie Côté, and Annie Taillon. 2010. Virtual reality induces dissociation and lowers sense of presence in objective reality. *Cyberpsychology, Behavior, and Social Networking* 13, 4 (2010), 429–435. https://doi.org/10.1089/cyber.2009.0164

[2] Jan Philipp Albrecht. 2016. How the GDPR will change the world. *Eur. Data Prot. L. Rev.* 2 (2016), 287.

[3] Dmitry Alexandrovsky, Susanne Putze, Michael Bonfert, Sebastian Höffner, Pitt Michelmann, Dirk Wenig, Rainer Malaka, and Jan David Smeddinck. 2020. Examining Design Choices of Questionnaires in VR User Studies. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–21. https://doi.org/10.1145/3313831.3376260

[4] Florian Alt, Stefan Schneegaß, Albrecht Schmidt, Jörg Müller, and Nemanja Memarovic. 2012. How to Evaluate Public Displays. In *Proceedings of the 2012 International Symposium on Pervasive Displays* (Porto, Portugal) *(PerDis '12)*. Association for Computing Machinery, New York, NY, USA, Article 17, 6 pages. https://doi.org/10.1145/2307798.2307815

[5] Florian Alt, Stefan Schneegass, Alireza Sahami Shirazi, Mariam Hassib, and Andreas Bulling. 2015. Graphical Passwords in the Wild: Understanding How Users Choose Pictures and Passwords in Image-Based Authentication Schemes. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Copenhagen, Denmark) *(MobileHCI '15)*. Association for Computing Machinery, New York, NY, USA, 316–322. https://doi.org/10.1145/2785830.2785882

[6] Morten Sieker Andreasen, Henrik Villemann Nielsen, Simon Ormholt Schrøder, and Jan Stage. 2007. What Happened to Remote Usability Testing? An Empirical Study of Three Methods. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '07)*. Association for Computing Machinery, New York, NY, USA, 1405–1414. https://doi.org/10.1145/1240624.1240838

[7] Roya Bagheri. 2016. Virtual Reality: The Real Life Consequences. *UC Davis Bus. LJ* 17 (2016), 101. https://heinonline.org/HOL/LandingPage?handle=hein.journals/ucdbulj17&div=7

[8] Louise Barkhuus and Jennifer A. Rode. 2007. From Mice to Men - 24 Years of Evaluation in CHI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '07)*. Association for Computing Machinery, New York, NY, USA, 1. https://doi.org/10.1145/1240624.2180963

[9] Evren Bozgeyikli, Andrew Raij, Srinivas Katkoori, and Rajiv Dubey. 2016. Point & Teleport Locomotion Technique for Virtual Reality. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play* (Austin, Texas, USA) *(CHI PLAY '16)*. Association for Computing Machinery, New York, NY, USA, 205–216. https://doi.org/10.1145/2967934.2968105

[10] Ian Bramley, Alastair Goode, Laura Anderson, and Elisabeth Mary. 2018. Researching in-store, at home: Using virtual reality within quantitative surveys. *International Journal of Market Research* 60, 4 (2018), 344–351.

[11] Daniel Buschek, Benjamin Bisinger, and Florian Alt. 2018. ResearchIME: A mobile keyboard application for studying free typing behaviour in the wild. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–14.

[12] Daniel Buschek, Fabian Hartmann, Emanuel Von Zezschwitz, Alexander De Luca, and Florian Alt. 2016. Snapapp: Reducing authentication overhead with a time-constrained fast unlock option. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 3736–3747.

[13] Kelly Caine. 2016. *Local Standards for Sample Size at CHI.* Association for Computing Machinery, New York, NY, USA, 981–992. https://doi.org/10.1145/2858036.2858498

[14] Sue V. G. Cobb, Sarah Nichols, Amanda Ramsey, and John R. Wilson. 1999. Virtual Reality-Induced Symptoms and Effects (VRISE). *Presence: Teleoper. Virtual Environ.* 8, 2 (April 1999), 169–186. https://doi.org/10.1162/105474699566152

[15] Karen Collins, Kamen Kanev, and Bill Kapralos. 2010. Using games as a method of evaluation of usability and user experience in human-computer interaction design. In *Proceedings of the 13th International Conference on Humans and Computers*. Citeseer, 5–10.

[16] Annabelle Cumyn, Kathleen Ouellet, Anne-Marie Côté, Caroline Francoeur, and Christina St-Onge. 2019. Role of researchers in the ethical conduct of research: A discourse analysis from different stakeholder perspectives. *Ethics & Behavior* 29, 8 (2019), 621–636.

[17] M. Danesi. 2006. Proxemics. In *Encyclopedia of Language & Linguistics (Second Edition)* (second edition ed.), Keith Brown (Ed.). Elsevier, Oxford, 241 – 243. https://doi.org/10.1016/B0-08-044854-2/01441-3

[18] Nigel Davies, Sarah Clinch, and Florian Alt. 2014. Pervasive displays: understanding the future of digital signage. *Synthesis Lectures on Mobile and Pervasive Computing* 8, 1 (2014), 1–128.

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

[20] Jaybie A. De Guzman, Kanchana Thilakarathna, and Aruna Seneviratne. 2019. Security and Privacy Approaches in Mixed Reality: A Literature Survey. *ACM Comput. Surv.* 52, 6, Article 110 (Oct. 2019), 37 pages. https://doi.org/10.1145/3359626

[21] Jing Du, Yangming Shi, Zhengbo Zou, and Dong Zhao. 2018. CoVR: Cloud-based multiuser virtual reality headset system for project communication of remote users. *Journal of Construction Engineering and Management* 144, 2 (2018), 04017109.

[22] Markus Funk, Florian Müller, Marco Fendrich, Megan Shene, Moritz Kolvenbach, Niclas Dobbertin, Sebastian Günther, and Max Mühlhäuser. 2019. Assessing the Accuracy of Point & Teleport Locomotion with Orientation Indication for Virtual Reality Using Curved Trajectories. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300377

[23] Laura Germine, Ken Nakayama, Bradley C Duchaine, Christopher F Chabris, Garga Chatterjee, and Jeremy B Wilmer. 2012. Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. *Psychonomic bulletin & review* 19, 5 (2012), 847–857.

[24] Michelle Goddard. 2017. The EU General Data Protection Regulation (GDPR): European regulation that has a global impact. *International Journal of Market Research* 59, 6 (2017), 703–705.

[25] Mattia Gustarini, Selim Ickin, and Katarzyna Wac. 2013. Evaluation of Challenges in Human Subject Studies "in-the-Wild" Using Subjects' Personal Smartphones. In *Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication* (Zurich, Switzerland) *(UbiComp '13 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 1447–1456. https://doi.org/10.1145/2494091.2496041

[26] Niels Henze and Susanne Boll. 2010. Push the Study to the App Store: Evaluating off-Screen Visualizations for Maps in the Android Market. In *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services* (Lisbon, Portugal) *(MobileHCI '10)*. Association for Computing Machinery, New York, NY, USA, 373–374. https://doi.org/10.1145/1851600.1851671

[27] Niels Henze, Martin Pielot, Benjamin Poppinga, Torben Schinke, and Susanne Boll. 2011. My app is an experiment: Experience from user studies in mobile app stores. *International Journal of Mobile Human Computer Interaction (IJMHCI)* 3, 4 (2011), 71–91.

[28] Niels Henze, Enrico Rukzio, and Susanne Boll. 2011. 100,000,000 Taps: Analysis and Improvement of Touch Performance in the Large. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services* (Stockholm, Sweden) *(MobileHCI '11)*. Association for Computing Machinery, New York, NY, USA, 133–142. https://doi.org/10.1145/2037373.2037395

[29] Marc Herz and Philipp A. Rauschnabel. 2019. Understanding the diffusion of virtual reality glasses: The role of media, fashion and technology. *Technological Forecasting and Social Change* 138 (2019), 228 – 242. https://doi.org/10.1016/j.techfore.2018.09.008

[30] Juan Luis Higuera-Trujillo, Juan López-Tarruella Maldonado, and Carmen Llinares Millán. 2017. Psychological and physiological human responses to simulated and real environments: A comparison between Photographs, 360 Panoramas, and Virtual Reality. *Applied Ergonomics* 65 (2017), 398–409. https://doi.org/10.1016/j.apergo.2017.05.006

[31] E. Hodgson, E. Bachmann, D. Waller, A. Bair, and A. Oberlin. 2012. Virtual reality in the wild: A self-contained and wearable simulation system. In *2012 IEEE Virtual Reality Workshops (VRW)*. 157–158.

[32] Bernd Huber and Krzysztof Z. Gajos. 2020. Conducting online virtual environment experiments with uncompensated, unsupervised samples. *PLOS ONE* 15, 1 (01 2020), 1–17. https://doi.org/10.1371/journal.pone.0227629

[33] Wijnand A IJsselsteijn, Yvonne AW de Kort, and Karolien Poels. 2013. The game experience questionnaire. *Eindhoven: Technische Universiteit Eindhoven* (2013), 3–9.

[34] Konrad Jamrozik. 2004. Research ethics paperwork: what is the plot we seem to have lost? *Bmj* 329, 7460 (2004), 286–287.

[35] Patricia Keith-Spiegel, Gerald P Koocher, and Barbara Tabachnick. 2006. What scientists want from their research ethics committee. *Journal of Empirical Research on Human Research Ethics* 1, 1 (2006), 67–81.

[36] Kay Köhle, Matthias Hoppe, Albrecht Schmidt, and Ville Mäkelä. 2021. Diegetic and Non-Diegetic Health Interfaces in VR Shooter Games. In *IFIP Conference on Human-Computer Interaction*. Springer.

[37] Edith L. M. Law, Luis Von Ahn, Roger B. Dannenberg, and Mike Crawford. [n.d.]. TAGATUNE: A GAME FOR MUSIC AND SOUND ANNOTATION.

[38] Xing Liu, Tianyu Chen, Feng Qian, Zhixiu Guo, Felix Xiaozhu Lin, Xiaofeng Wang, and Kai Chen. 2017. Characterizing Smartwatch Usage in the Wild. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services* (Niagara Falls, New York, USA) *(MobiSys '17)*. Association for Computing Machinery, New York, NY, USA, 385–398. https://doi.org/10.1145/3081333.3081351

[39] Xiao Ma, Megan Cackett, Leslie Park, Eric Chien, and Mor Naaman. 2018. Web-Based VR Experiments Powered by the Crowd. In *Proceedings of the 2018 World Wide Web Conference* (Lyon, France) *(WWW '18)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 33–43. https://doi.org/10.1145/3178876.3186034

[40] Christian Mai, Tim Wiltzius, Florian Alt, and Heinrich Hußmann. 2018. Feeling Alone in Public: Investigating the Influence of Spatial Layout on Users' VR Experience. In *Proceedings of the 10th Nordic Conference on Human-Computer Interaction* (Oslo, Norway) *(NordiCHI '18)*. Association for Computing Machinery, New York, NY, USA, 286–298. https://doi.org/10.1145/3240167.3240200 mai2018nordichi.

[41] Ville Mäkelä, Rivu Radiah, Saleh Alsherif, Mohamed Khamis, Chong Xiao, Lisa Borchert, Albrecht Schmidt, and Florian Alt. 2020. Virtual Field Studies: Conducting Studies on Public Displays in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. ACM, New York, NY, USA. https://doi.org/10.1145/3313831.3376796

[42] Ville Mäkelä, Sumita Sharma, Jaakko Hakulinen, Tomi Heimonen, and Markku Turunen. 2017. Challenges in Public Display Deployments: A Taxonomy of External Factors. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 3426–3475. https://doi.org/10.1145/3025453.3025798

[43] Guido Makransky, Richard E Mayer, Nicola Veitch, Michelle Hood, Karl Bang Christensen, and Helen Gadegaard. 2019. Equivalence of using a desktop virtual reality science simulation at home and in class. *Plos one* 14, 4 (2019), e0214944.

[44] Nemanja Memarovic, Marc Langheinrich, Keith Cheverst, Nick Taylor, and Florian Alt. 2013. P-LAYERS – A Layered Framework Addressing the Multifaceted Issues Facing Community-Supporting Public Display Deployments. *ACM Trans. Comput.-Hum. Interact.* 20, 3, Article 17 (July 2013), 34 pages. https://doi.org/10.1145/2491500.2491505

[45] RJ Menzies, SJ Rogers, AM Phillips, Elodie Chiarovano, C De Waele, Frans AJ Verstraten, and H MacDougall. 2016. An objective measure for the visual fidelity of virtual reality and the risks of falls in a virtual environment. *Virtual Reality* 20, 3 (2016), 173–181. https://doi.org/10.1007/s10055-016-0288-6

[46] Aske Mottelson and Kasper Hornbæk. 2017. Virtual Reality Studies Outside the Laboratory. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology* (Gothenburg, Sweden) *(VRST '17)*. Association for Computing Machinery, New York, NY, USA, Article 9, 10 pages. https://doi.org/10.1145/3139131.3139141

[47] Fares Moustafa and Anthony Steed. 2018. A Longitudinal Study of Small Group Interaction in Social Virtual Reality. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology* (Tokyo, Japan) *(VRST '18)*. Association for Computing Machinery, New York, NY, USA, Article 22, 10 pages. https://doi.org/10.1145/3281505.3281527

[48] T. C. Peck, L. E. Sockol, and S. M. Hancock. 2020. Mind the Gap: The Underrepresentation of Female Participants and Authors in Virtual Reality Research. *IEEE Transactions on Visualization and Computer Graphics* 26, 5 (2020), 1945–1954. https://doi.org/10.1109/TVCG.2020.2973498

[49] Radiah Rivu, Yumeng Zhou, Robin Welsch, Ville Mäkelä, and Florian Alt. 2021. When Friends become Strangers: Understanding the Influence of Avatar Gender On Interpersonal Distance in Virtual Reality. In *IFIP Conference on Human-Computer Interaction*. Springer.

[50] David Saffo, Sara Di Bartolomeo, Caglar Yildirim, and Cody Dunne. 2021. Remote and Collaborative Virtual Reality Experiments via Social VR Platforms. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411764.3445426

[51] David Saffo, Caglar Yildirim, Sara Di Bartolomeo, and Cody Dunne. 2020. Crowdsourcing Virtual Reality Experiments Using VRChat. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3334480.3382829

[52] Alireza Sahami Shirazi, Niels Henze, Tilman Dingler, Martin Pielot, Dominik Weber, and Albrecht Schmidt. 2014. Large-Scale Assessment of Mobile Notifications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 3055–3064. https://doi.org/10.1145/2556288.2557189

[53] Stefan Schneegass, Frank Steimle, Andreas Bulling, Florian Alt, and Albrecht Schmidt. 2014. SmudgeSafe: Geometric Image Transformations for Smudge-Resistant User Authentication. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Seattle, Washington) *(UbiComp '14)*. Association for Computing Machinery, New York, NY, USA, 775–786. https://doi.org/10.1145/2632048.2636090

[54] Valentin Schwind, Pascal Knierim, Nico Haas, and Niels Henze. 2019. Using presence questionnaires in virtual reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[55] Mel Slater, John McCarthy, and Francesco Maringelli. 1998. The Influence of Body Movement on Subjective Presence in Virtual Environments. *Human Factors* 40, 3 (1998), 469–477. https://doi.org/10.1518/001872098779591368 arXiv:https://doi.org/10.1518/001872098779591368 PMID: 9849105.

[56] Mel Slater and Anthony Steed. 2000. A Virtual Presence Counter. *Presence: Teleoper. Virtual Environ.* 9, 5 (Oct. 2000), 413–434. https://doi.org/10.1162/105474600566925

[57] James S Spiegel. 2018. The ethics of virtual reality technology: social hazards and public policy recommendations. *Science and engineering ethics* 24, 5 (2018), 1537–1550. https://doi.org/10.1007/s11948-017-9979-y

[58] A. Steed, S. Frlston, M. M. Lopez, J. Drummond, Y. Pan, and D. Swapp. 2016. An 'In the Wild' Experiment on Presence and Embodiment using Consumer Virtual Reality Equipment. *IEEE Transactions on Visualization and Computer Graphics* 22, 4 (2016), 1406–1414.

[59] Anthony Steed, Francisco R Ortega, Adam S Williams, Ernst Kruijff, Wolfgang Stuerzlinger, Anil Ufuk Batmaz, Andrea Stevenson Won, Evan Suma Rosenberg, Adalberto L Simeone, and Aleshia Hayes. 2020. *Evaluating immersive experiences during Covid-19 and beyond.* https://interactions.acm.org/blog/view/evaluating-immersive-experiences-during-covid-19-and-beyond

[60] The Against Gravity Team. 2019. *Rec Room in 2018.* https://recroom.com/2018inreview/

[61] Douglas Turnbull, Ruoran Liu, Luke Barrington, and Gert Lanckriet. 2007. A game-based approach for collecting semantic annotations of music. In *In 8th International Conference on Music Information Retrieval (ISMIR.*

[62] 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/Addison-Wesley Publishing Co., USA, 359–364. https://doi.org/10.1145/311535.311589

[63] Luis von Ahn and Laura Dabbish. 2004. Labeling Images with a Computer Game. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vienna, Austria) *(CHI '04).* Association for Computing Machinery, New York, NY, USA, 319–326. https://doi.org/10.1145/985692.985733

[64] FengJun Zhang, GuoZhong Dai, and Xiaolan PENG. 2016. A survey on human-computer interaction in virtual reality. *Scientia Sinica Informationis* 46, 12 (2016), 1711–1736.