Technische Universität München
TUM School of Engineering and Design

Revisiting Street Crossing Simulators—New Approaches in Virtual and Augmented Reality

Philipp Bastian Maruhn

Vollständiger Abdruck der von der TUM School of Engineering and Design der Technischen Universität München zur Erlangung des akademischen Grades eines Doktors der Ingenieurwissenschaften (Dr.-Ing.) genehmigten Dissertation.

Vorsitz:	Prof. Dr.-Ing. Birgit Vogel-Heuser
Prüfende der Dissertation:	1. Prof. Dr. phil. Klaus Bengler 2. Prof. Gudrun J. Klinker, Ph.D.

Die Dissertation wurde am 22.04.2022 bei der Technischen Universität München eingereicht und durch die TUM School of Engineering and Design am 07.10.2022 angenommen.

Danke

Die Möglichkeit zu promovieren ist alles andere als eine Selbstverständlichkeit, weshalb ich die Gelegenheit nutzen möchte mich an dieser Stelle zu bedanken. Danke an alle, die mir diese Erfahrung ermöglicht und so einzigartig gemacht haben. Zunächst richtet sich mein Dank an meine betreuende Institution, die Technische Universität München und den Lehrstuhl für Ergonomie. Unterstützt durch Mittel der Deutschen Forschungsgemeinschaft wurde mir hier ein wunderbares Umfeld geboten, in dem sich meine Forschung frei entfalten durfte. Ein Umstand, dem man nur allzu leicht eine zu geringe Wertschätzung entgegenbringt, welchem ich aber nicht dankbar genug sein kann. In diesem Zusammenhang gilt mein Dank auch meinem Betreuer, Prof. Klaus Bengler und meiner Zweitgutachterin, Prof. Gudrun Klinker.

Keine Selbstverständlichkeit ist auch das kollegiale, nein, freundschaftliche Miteinander am Lehrstuhl für Ergonomie. Auch wenn Garching an manchen Tagen weiter weg erschien, habe ich mich immer gefreut an den Lehrstuhl zu kommen. Vor allem habe ich genossen immer Gleichgesinnte zu finden, die sich für wissenschaftliche und manchmal auch nicht ganz so wissenschaftliche Ideen begeistern lassen.

Danke an Sonja Schneider und Cornelia Gruber für eure Geduld und Ausdauer bei all meinen Statistikfragen. Danke Martin Fleischer für jedes "brauchst du Hilfe", wenn ich an deiner Tür vorbeigegangen bin. Danke Annika Boos für die tiefgründigen Diskussionen. Danke André Dietrich für deinen Kickstart im Fußgängersimulator, die vielen gemeinsamen Stunden in Unity und, zusammen mit Olivia Herzog, das beste VR Seminar. Danke Ingrid Bubb, Anna Feldhütter und Jonas Radlmayr für jedes Blatt das wir (nicht) gespielt haben. Danke Lorenz Prasch für die Zeit in Makerspace, DMZ, LaTeX, Illustrator, 3301, VR Garage, jeden Kommentar zu dieser Arbeit und alles was hier keinen Platz mehr gefunden hat.

Danke an meine Familie, die mir meinen Weg hierher ermöglicht hat. Danke Moni, dass du mich immer unterstützt und jeden Tag ein Lachen ins Gesicht zauberst.

Abstract

Virtual reality has been used as a research tool for several decades and is well established in the study of pedestrian behavior. Road crossings in particular represent a safety-critical process and are often investigated using virtual environments. In particular, virtual reality headsets have increasingly been used in this regard. Many open questions remain, however. How should these tools ideally be applied? What limitations do they currently face? And how can these obstacles be overcome in the future? This work highlights different aspects of these concerns. For example, a realistic distance estimation is crucial for pedestrian simulators; however, distances are generally underestimated in virtual environments. Article 1 investigates the relationship between the mode of locomotion and different protocols for measuring distance perception. The findings indicate that natural walking, compared with a solely stationary experience, can improve distance perception. Article 2 demonstrates that the use of a virtual representation of one's own body, or avatar, influences crossing behavior in pedestrian simulators. Participants accepted smaller inter-vehicular gaps to cross a street when an avatar was displayed than without one. The following papers investigate the suitability of two additional virtual reality technologies, namely augmented reality (Article 3) and a low-cost variant for at-home testing (Article 4). This hardware comparison reveals that both approaches are suitable for virtual pedestrian studies. Additionally, it can be confirmed that experiments—when properly designed—can be conducted alone at home without a supervisor. In this scenario, the participants in the remote setting rated their behavior being more realistic than the ones in the laboratory (Article 4). However, across all studies, neither the new nor the established approaches show absolute validity.

Contents

Glossary

5G	Fifth Generation Cellular Network Technology
AR	Augmented Reality
AV	Augmented Virtuality
CAVE	Cave Automatic Virtual Environment
CCTV	Closed Circuit Television
CS	Cybersickness
FoV	Field of View
HMD	Head-Mounted Display
IMU	Inertial Measurement Unit
NPC	Non-Player Character
RDW	Redirected Walking
VE	Virtual Environment
VR	Virtual Reality

1

Introduction

1.1 A Brief History of Virtual Reality[1]

People have been on a quest to escape reality and immerse themselves in a world of fiction for much of human history. Technological progress has merely changed how they do so. Arts, in the form of storytelling, performances, and paintings, have accompanied humanity throughout. Technical innovations have found new ways first to present a fictional world and second to give the audience a feeling of being a part of this world. Slater (2003) distinguishes between these two processes as immersion and presence. Immersion simply describes “what the technology delivers from an objective point of view” (Slater, 2003, p. 1), whereas presence is the individual “human reaction to immersion” (Slater, 2003, p. 2). Toward the end of the 18th century, for example, 360-degree panorama paintings represented a novel approach to increasing immersion. Visitors entered the large rotundas that housed these paintings and subsequently found themselves in the middle of the scenery (Bown et al., 2017). Around that time, Queen Charlotte of Great Britain and Ireland visited a panorama showing a sea scene and said she would become seasick (Altick, 1978). On the one hand, this indicates a high sense of presence; on the other hand, it can also be described as a predecessor of simulator sickness (Bown et al., 2017). At the same time, analog antecedents of today’s

[1]This sections narrative is inspired by the great historical overview on virtual reality by Bown et al. (2017).

virtual reality (VR) glasses were being developed. In 1838, Charles Wheatstone described the first apparatus that made it possible to simultaneously watch a slightly offset view of the same scene with each eye, thus creating an illusion of depth: the stereoscope.

Based on this large and rather cumbersome apparatus, David Brewster developed the predecessor of today's head-mounted displays (HMDs). By combining the principles of stereoscopy with magnifying lenses and prisms, Brewster (1856) created the first handheld stereoscopic viewer, the lenticular stereoscope (see Figure 1.1). With the development of photography and stereo-photography, stereoscopes quickly grew in popularity. They can be considered "a 19th-century version of the first cheap, take-home, VR system" (Bown et al., 2017, p. 245). The same technology is still used today in trademarked View-Master devices and has remained largely unchanged for nearly a century, even though the latest devices use smartphones instead of rotating discs to display stereo images.

It would take until the advent of the information age for inventors to take the first steps in the direction of VR in the modern sense. With his Stereoscopic-Television Apparatus for Individual Use (Heilig, 1957), also referred to as the Telesphere Mask, and the Sensorama (Heilig, 1961), the American cinematographer Morton Heilig created a truly immersive experience. The Telesphere Mask already has an unmistakable resemblance to today's headsets (Figure 1.2). In the larger machine known as the Sensorama (see Figure 1.3), the visitor could experience one of five 3D films. Aside from stereoscopy, the films were synchronized with odor generators, vibrating seats, and a wind machine. In addition to their application in the entertainment industry, Heilig already realized the potential of VR devices elsewhere: "There are increasing demands today for ways and means to teach and train individuals without actually subjecting the individuals to possible hazards of particular situations" (Heilig, 1961, p. 9).

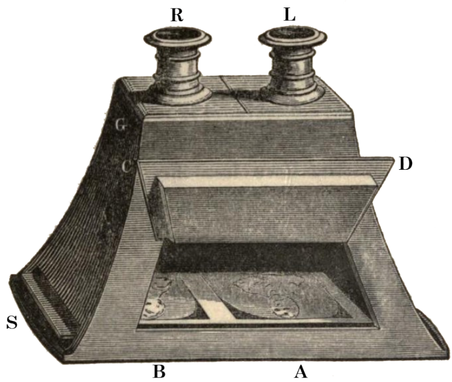

Figure 1.1: Stereoscope by Brewster (1856). A stereo image pair (A,B) is inserted through the opening S. Transparent images are illuminated through a glass plate at the bottom. A hatch (C,D) can be opened to allow light to shine on opaque images. The image pair is viewed through the left and right eye apertures (L,R) in conjunction with the lenses (G), giving the impression of depth.

Although the term "virtual reality" emerged decades later, it will be briefly described here. Many definitions exist for VR (Mazuryk & Gervautz, 1996; Zhou & Deng, 2009). The spectrum ranges from hardware-specific descriptions of an apparatus and its components through to philosophical discussions. Fuchs et al. (2019, p. 8) provide a technical definition of VR as "a scientific and technical domain that uses computer science (1) and behavioural interfaces (2) to simulate in a virtual world (3) the behaviour of 3D entities, which interact in real time (4) with each other and with one or more users in pseudo-natural immersion (5) via sensorimotor channels." What this definition in particular and all the others have in common, though, is that some kind of interaction between the user and the virtual environment (VE) is assumed.

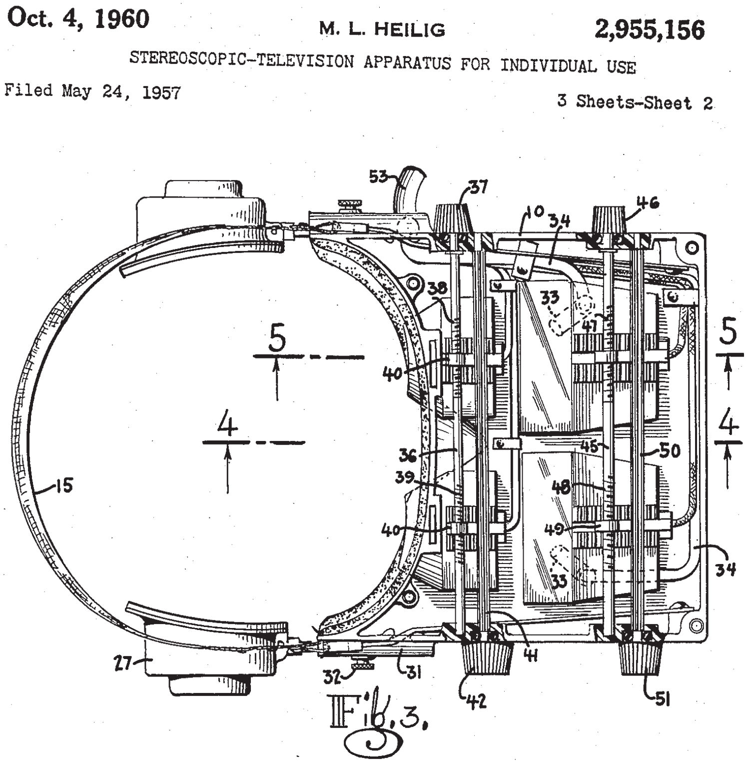

Figure 1.2: Stereoscopic-television apparatus by Heilig (1957).

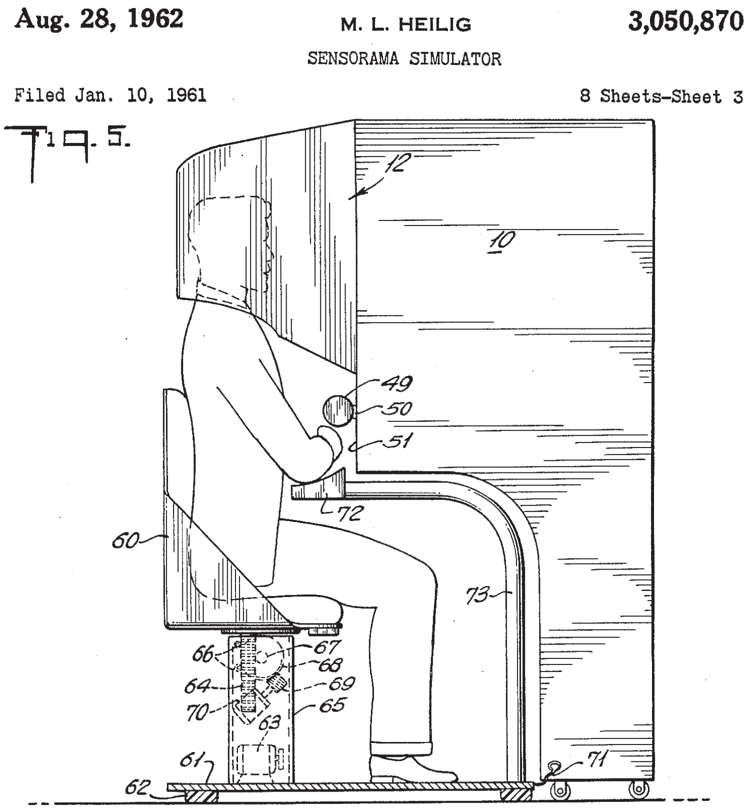

Figure 1.3: Sensorama by Heilig (1961).

The devices presented so far indeed made it possible to experience a three-dimensional virtual world, but not to interact with it. Further development was necessary in pursuit of Ivan Sutherland's vision of an "ultimate display":

> "The ultimate display would, of course, be a room within which the computer can control the existence of matter. A chair displayed in such a room would be good enough to sit in. Handcuffs displayed in such a room would be confining, and a bullet displayed in such a room would be fatal. With appropriate programming such a display could literally be the Wonderland into which Alice walked.
>
> (Sutherland, 1965, p. 508)

Comeau and Bryan (1961) were the first to incorporate head tracking and thus a way of interacting with the displayed content. A helmet tracked the head's rotation, which was linked to a closed circuit television (CCTV) camera. The camera's live stream was then played back inside the helmet (Bown et al., 2017; Rid, 2016). Inspired by this setup (Sutherland & Sproull, 1996), Sutherland created what is known today as the first HMD. Instead of streaming a video image, Sutherland (1968) displayed a computer-generated image that was updated according to the HMD's position. For the first time, "people could observe where they were in a synthetic world" (Sutherland & Sproull, 1996, 00:49:35).

Even back then, the question arose as to whether the headset itself would determine the position in the room (inside-out tracking) or whether external sensors would be used for this purpose (outside-in tracking). Sutherland (1968) presented two different approaches to track the HMD's position. The first involved three ultrasonic transmitters attached to the HMD. Four receivers were mounted in a square on the ceiling and received these signals. However, this ultrasound-based "showerhead" setup was susceptible to interference, for example from air turbulence from the air-conditioning system (Sutherland & Sproull, 1996). The second approach pursued a mechanical solution: A mechanical linkage to the ceiling allowed the head position to be reliably tracked along six degrees of freedom. Due to the long pole from headset to ceiling, this setup is known today as "the Sword of Damocles" and often referred to as the first VR headset. However, concerning its properties, it should rather be considered the first augmented reality (AR) headset. Two cathode ray tubes, one for each eye, generated a virtual image, which was then presented to the user via semitransparent prisms in front of the eyes (Sutherland, 1968). Thus, the user could see the virtual object, apparently floating in the real room.

While the basic principles were hence proven to work, it took several years for computer graphics to catch up. Starting in 1984, the Virtual Visual Environment Display (VIVED) system, "the first low-cost, wide field-of-view, stereo, head-tracked, head-mounted display" (McGreevy, 1993, p. 165) was developed at NASA. Around the same time, Jaron Lanier founded Visual Programming Languages, manufactured the first commercially available HMDs, and introduced the term "virtual reality" (Bown et al., 2017). From this process of development, mainly driven by continuous improvements to display technologies, today's VR glasses would eventually emerge.

In parallel with HMDs, another method to experience VEs immersively emerged in the 1990s. At that time, HMDs faced several shortcomings, mainly regarding their limited field of view (FoV) and low display resolution. Additionally, the devices were limited to one user each. In 1992, Cruz-Neira et al. first introduced a projection-based system, the Cave Automatic Virtual Environment (CAVE) (Cruz-Neira et al., 1992; McLellan, 1996). A CAVE in this context is a room where the walls and sometimes the floor and ceiling consist of rear-projection surfaces. The user is equipped with shutter glasses to experience the projection as three-dimensional. Additionally, the position of the user's head is tracked to update the projected view according to their perspective (Cruz-Neira et al., 1993). This implementation also results in the user being able to see their own body in the VE, unlike in an HMD. CAVEs are nowadays used in a variety of different applications, such as "in military and training applications, education, medicine, scientific visualizations, and many other areas of human activity" (Muhanna, 2015, p. 355). However, such systems are always associated with high financial costs due to the complexity of the setups.

The past decade in particular has seen a leap in the development of HMDs. Disappointed with the performance of headsets available at the time, Palmer Luckey founded OculusVR in 2012 and developed the first commercially successful VR headset for gaming (Bown et al., 2017; Kumparak, 2014). Other manufacturers would soon follow suit, starting a second wave of VR HMDs. With its Lighthouse Tracking System, Valve (a software developer company) in cooperation with the consumer electronics manufacturer HTC released the first consumer room-scale tracking technology. This enabled a larger number of researchers to use VR as a research tool, although some inaccuracies had to be accepted under certain circumstances, mainly connected to tracking losses (Niehorster et al., 2017). Otherwise, the degree of accuracy seemed sufficient to analyze human behavior in VR (Verdelet et al., 2019).

Technical innovations have also allowed for new variations of HMDs. The dissemination of powerful smartphones featuring high-resolution displays has enabled a new category of HMD known as mobile or phone-based VR. The phone's inertial measurement unit (IMU) is used to track head rotation. Similar to the 19th-century stereoscopes described above, the phone is inserted into a holder, and lenses magnify the image. The display screen is divided into two parts, one for each eye. Devices range from high-end, cushioned viewers to low-cost solutions such as the Google Cardboard (Smus et al., 2014). Aside from HMDs that rely on an external processing unit and phone-based VR, another category of HMDs has emerged: wireless, standalone headsets. In 2016, Microsoft released the HoloLens, a standalone AR headset. However, the limited FoV reduces the sense of immersion (Avila & Bailey, 2016) and thus also limits the range of use. With the Oculus Quest gaming system, Facebook released a standalone VR headset with inside-out tracking in 2019 that potentially enables unlimited tracking space (OculusVR, 2019).

A look at the history of VR shows that technological achievements are constantly creating new ways to present virtual worlds in an ever more immersive way. Old ideas are repeatedly combined with new concepts. However, the goal of creating the "ultimate display" is still being pursued. In recent years, development has increasingly accelerated, and the range of applications for VR has drastically expanded. Due to its increasing cost-effectiveness and wide availability, VR is being used more frequently as a tool in various research fields, especially those investigating human factors. Simulators in which human behavior in road traffic is studied benefit particularly from further developments in the field of VR. In addition to driving simulators, the number of simulators that investigate behavior from the perspective of a pedestrian is also on the rise.

1.2 Pedestrians from a Human Behavioral Perspective

The number of annual traffic fatalities reached 1.35 million in 2016 (World Health Organization, 2018). Compared with the growing world population and increasing motorization, this does not represent a relative increase in the number of accidents, however, no decrease can be observed either (World Health Organization, 2018). The UN sustainable development goal of halving the number of traffic fatalities by 2020 (UN General Assembly, 2015) could not be met and is therefore still relevant.

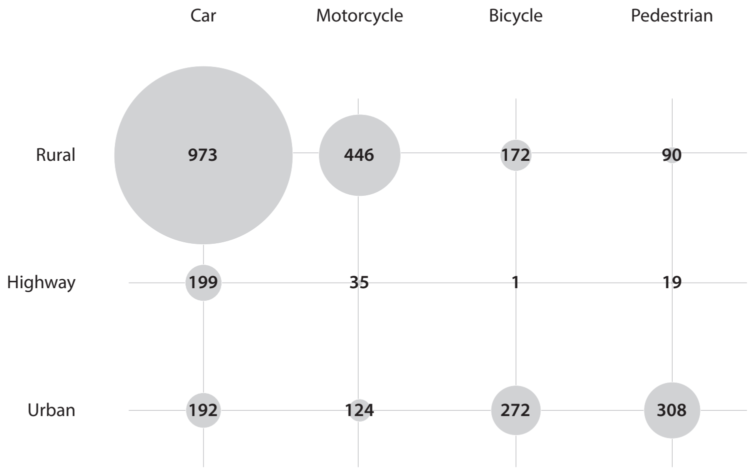

Figure 1.4: Traffic fatalities in 2019 in Germany by mode of transportation and location (Statistisches Bundesamt, 2020b).

Pedestrians account for 23% of all traffic fatalities worldwide (World Health Organization, 2018). Depending on the mode of transportation and the environment, the risk of accidents varies. For example, the risk of suffering a fatal road traffic accident in Germany in 2019 was third highest for pedestrians in urban areas (see Figure 1.4). In general, crossing the street seems to be particularly dangerous.

> “ Most pedestrian collisions happen when pedestrians are crossing the road, rather than walking or standing alongside the road.
>
> (World Health Organization, 2013, p. 4)

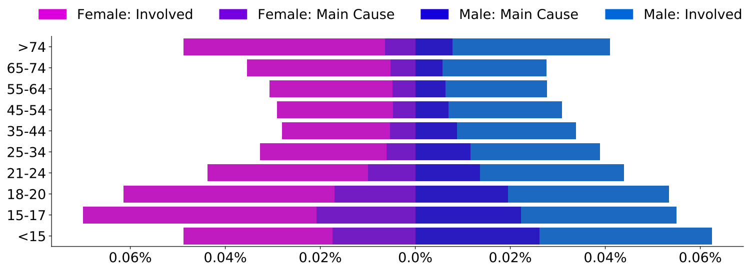

Figure 1.5: Pedestrian accidents by age and gender in Germany 2019 as proportion of the total population. Accident data retrieved from Statistisches Bundesamt (2021b); population data retrieved from Statistisches Bundesamt (2021a).

The International Ergonomics Association (2021, section: Definition and Applications) defines ergonomics, often referred to as human factors, as "the scientific discipline concerned with the understanding of interactions among humans and other elements of a system, and the profession that applies theory, principles, data, and methods to design in order to optimize human well-being and overall system performance."

Human factors in transportation systems and road traffic are one area of research. In the case of pedestrians, for example, the question arises as to what the reasons are for the high number of accidents and how they can be prevented. To achieve the latter, the reasons for traffic accidents must first be identified.

The pedestrian accident statistics for Germany (see Statistisches Bundesamt, 2020a, p. 304) suggest that pedestrians not only are frequently involved in accidents but also often cause them: In 2019 alone, pedestrian misconduct caused 13,065 accidents resulting in personal injury. Alcohol ($N = 591$) or drug influence ($N = 43$) played an almost negligible role. By far the most common cause of accidents ($N = 10,004$) was wrong behavior while crossing the street. Of these incidents, the most common reason was pedestrians not paying attention to vehicle traffic ($N = 5,785$), followed by suddenly stepping out from behind visual obstacles ($N = 1,799$). The pedestrian accident figures (see Figure 1.5), also highlight an age effect: Children and young adults especially, as well as older pedestrians, tend to be more likely to be involved in or cause accidents.

> "Chronological age itself, however, does not explain why older adults are more vulnerable at the roadside; rather, a deterioration or difficulty in one or more of the processes needed to cross the road safely could result in an increase in rate of injury. In order to successfully execute a road crossing, pedestrians must perceive and pay attention to vehicles approaching from both directions. They need to detect approaching traffic, determine the velocity of approaching vehicles, and estimate if they have enough time to cross before the approaching vehicle reaches their crossing path. Once the decision to cross has been taken, pedestrians must execute a crossing movement, reevaluating the risks as they go.
>
> (Wilmut & Purcell, 2021, pp. 1-2)

Wilmut and Purcell (2021) conducted a literature survey to identify risk factors when elderly pedestrians cross the street. As described in their paper (see quote above), crossing the street is a complex task. The information processing model (see Figure 1.6) developed by Wickens and Flach (1988) can be applied to categorize the individual sub-tasks of a street crossing as well as possible influencing factors.

Age-related impairments could play a role, for example, in sensing traffic and the surrounding scenery. After sensory detection, a perception process follows. Age-related effects could also play a role here—in this case, for younger road users. Based on self-reports, Connelly et al. (1998) found that 5–9-year-old pedestrians mainly rely on the distance of a vehicle to decide whether it is safe to cross. However, this leads to an increased risk with high vehicle speeds when the time

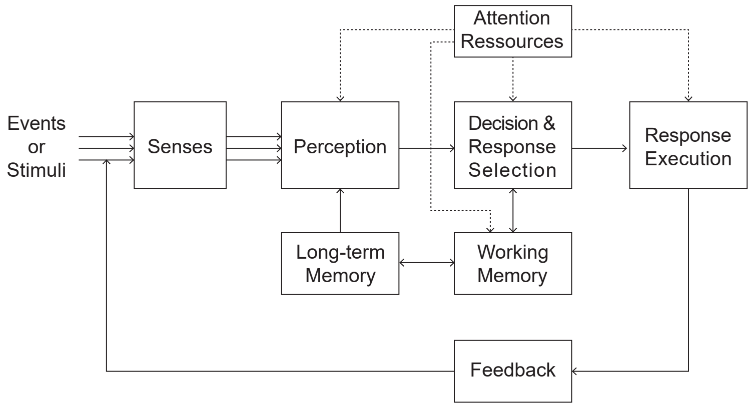

Figure 1.6: Information processing model based on Wickens and Carswell (2021) and Wickens and Flach (1988).

for a crossing becomes shorter while the gap size remains the same. In a dynamic system like road traffic, fast and efficient decision and response selection is crucial (Schwebel et al., 2012). However, young children in particular seem to take longer to make a crossing response selection than adults (Plumert et al., 2004; Schwebel et al., 2012; te Velde et al., 2005). During the response execution (i.e., crossing the street), elderly pedestrians could again be at greater risk due to slower walking speeds compared to younger adults (Wilmut & Purcell, 2021). Misdirected attention resources constitute an additional risk factor, for example distraction from smartphones while crossing the street (Horberry et al., 2019) or children running after an object or person (Schwebel et al., 2012).

The underlying causes of risky behavior are manifold, often controversial (Wilmut & Purcell, 2021), and not yet fully understood (Barton & Schwebel, 2007). Road traffic is also subject to continuous change. Increasing urbanization and vehicle automation pose new challenges for pedestrians (Bengler et al., 2018). With autonomous vehicles on the horizon, the additional question arises of how pedestrians will interact with these driverless vehicles in the future (Millard-Ball, 2018). Human crossing behavior must be understood in general, to properly precondition autonomous vehicles to expect certain human behavior. It is important to examine these and future research questions from the perspective of ergonomics. In recent years, a mix of methods has been established for investigating pedestrian behavior. In addition to traffic observations or studies on a test track, methods also include the increasing use of virtual environments.

1.3 Pedestrians on the Reality-Virtuality Continuum

So far, the content described in this thesis has been limited to VR, with the exception of the Sword of Damocles, which is an AR headset (see Section 1.1). However, between reality on the one hand and virtual reality on the other, there is a wide range of possibilities for combining virtual and real content. Milgram defines this area between reality and virtual reality as the "reality-virtuality continuum", or "mixed reality" (Milgram & Kishino, 1994; Milgram et al., 1995). Milgram et al. (1995) further divide this area into augmented reality and augmented virtuality (AV), as listed in Table 1.1. In this context, reality describes an environment that is composed only of real objects and is perceived either in person or via video (Milgram et al., 1995). AR then describes a setting in which virtual objects are superimposed in three dimensions onto this otherwise solely real environment. Moving further along the scale, VR describes a completely virtual environment, while AV represents a virtual environment that is enriched with real stimuli.

Without noticing, we are confronted with AR and AV almost every day. For example, a viewer experiences AR when a sports broadcast is enriched with virtual information (Azuma, 1997), or AV when a real actor is situated among virtual scenery via green-screen technology.

Table 1.1: Possible measurement methods for the investigation of pedestrian behavior classified by the reality-virtuality continuum according to Milgram et al. (1995).

Reality	**Augmented Reality**	**Augmented Virtuality**	**Virtual Reality**
Experiments	AR Glasses	CAVE	PC-based
Observations and Surveys	Video Pass-Through	Kiosk	Standalone
Accident Databases			Mobile

As described in Section 1.1, technological progress is enabling these concepts to be introduced into an increasing number of domains. In the following paragraphs, this will be discussed in the context of research into pedestrian behavior.

> “ Basically, if the goal is to study properties of some natural setting itself, an experiment is not particularly appropriate. For studying theories that have abstracted some properties of natural settings, however experiments can be ideally suited.
>
> (Webster & Sell, 2014, p. 12)

To examine the behavior of pedestrians in different situations, different methods can be applied (see Table 1.1), depending on the underlying research question. The decisive factor is whether pedestrian behavior is to be studied holistically in its natural environment or whether individual aspects are to be investigated selectively. Depending on the situation, unobtrusive methods, such as traffic observations, or a more direct approach, such as laboratory experiments in controlled environments, is then applied (Webster & Sell, 2014, see quote above). Next, the tools and methods that are currently used, as well as what they are particularly suited for, are briefly introduced.

Accident databases help to identify potential risk areas. The accidents are mostly collected in national databases and forwarded to international institutions to enable cross-country comparisons (see Section 1.2). However, these national and international databases and accident reports are usually limited to meta-information (e.g., accident participants and location) and thus often do not allow conclusions to be drawn about the underlying causes of accidents. The aim of ergonomics is to understand human behavior, in this case the behavior of pedestrians, to avoid future accidents. Databases can provide information on where and how accidents happened (e.g., when crossing the street), but the question of "why" often remains unanswered. Another limitation is the retrospective nature of accident databases, whereby existing problems can be identified only after they have caused an accident at least once.

Observational studies are employed for a variety of different research questions to analyze pedestrian behavior (Mamidipalli et al., 2015). Examples described by Wilmut and Purcell (2021) include the following: recording walking speeds when crossing the road (Avineri et al., 2012; Hoxie & Rubenstein, 1994); behavior and compliance when crossing at designated crossings (Brosseau et al., 2013; Cloutier et al., 2017); and acceptance of gap sizes at unregulated road crossings (Harrell & Bereska, 1992; Naser et al., 2017). In these cases, the pedestrians are usually observed at several previously determined locations by one or two observers, or they are recorded with video cameras (Mamidipalli et al., 2015; Wilmut & Purcell, 2021). Occasionally, the observations are followed by ad hoc short interviews (Wilmut & Purcell, 2021).

As with the accident databases described above, knowledge can only be obtained from the road traffic that is currently present. However, this also means that the influence of technical or regulatory changes can only be observed after their introduction into road traffic. For example, it is not possible to study the everyday interaction of pedestrians and AVs, as this would require real or, to be more versatile, virtual prototypes applied in an experimental setting.

Conducting and analyzing observational studies is time-consuming. In addition, the desired event to be observed may occur very rarely, so even more observations are needed. It is therefore unsurprising that these studies are becoming increasingly automated (e.g., the incorporation of sensors to measure walking speeds or automated video annotation and analysis) (Mamidipalli et al., 2015). However, if specific aspects are to be studied in isolation, there is a limit to the usefulness of observations. Instead, these scenarios are investigated in dedicated experimental settings.

Real-world experiments are mainly carried out on closed roads (Rodríguez Palmeiro et al., 2018), on university campuses (Lundgren et al., 2017; Rodríguez Palmeiro et al., 2018), on dedicated test tracks (Faas, Mathis, et al., 2020), or on public roads (Connelly et al., 1998). The overarching research question is usually whether or not a pedestrian would cross the street in a specific scenario. However, in most study designs, participants are not actually instructed to cross the street due to safety concerns. In cases where subjects are asked to actually cross the road, their safety must be ensured with great effort, for example by using professional test track drivers within a closed-off area (Faas, Mathis, et al., 2020). Otherwise, the crossing intention can only be measured indirectly or be simulated. Different techniques are used for this. Walker et al. (2019) propose a handheld slider to measure on a continuous scale the pedestrian's willingness over time to cross the road. More often, however, subjects are asked to signal the time at which they would initiate their street crossing. This is usually done either by pressing a button or by taking a step toward the curb, with the latter reflecting more natural behavior (Faas, Mattes, et al., 2020).

However, experiments in a real setting are associated with some disadvantages. They typically involve a high level of organizational and financial effort. In addition to finding a suitable location, vehicles and additional personnel are usually required. Investigations are also carried out in real road traffic. Here, however, the control of the experimental condition is limited.

Pedestrian simulators, often referred to as street-crossing simulators (Cavallo et al., 2019), were first introduced as a methodological complement to overcome the shortcomings of real-world tests as described above. Real world objects (e.g., scenery and vehicles) are replaced by virtual objects to create a safe and controllable environment. This involves combining real and virtual objects in different ways (see Table 1.1). As described above, enriching a real environment with virtual content is known as augmented reality, or AR. Milgram et al. (1995) distinguish between monitor-based ("window on the world") and see-through AR displays.

Milgram et al. (1995) further divide see-through displays into optical see-through and video see-through types. In optical systems, the virtual content is superimposed on semitransparent displays. With video see-through systems, the user only sees a live video stream of the real environment, which is then augmented with virtual content. Recently, optical see-through systems have been incorporated to analyze pedestrian behavior based on the Microsoft HoloLens HMD (Hartmann et al., 2018; Perez et al., 2019). In contrast to VR approaches, only the objects of interest need to be virtually created while making use of the real environment (Hartmann et al., 2018). However, one major drawback of the currently available optical see-through AR headsets is their limited FoV, with recent models only achieving up to 52° diagonally (Xiong et al., 2020). To overcome this limitation, while still providing the benefit from using a real environment in combination with virtual vehicles, a video see-through approach has been developed in this thesis (see Article 3). Compared with the large number of virtual approaches, such an implementation constitutes an exception.

Virtual paradigms for pedestrian simulators can be divided into AV and VR (see Table 1.1). An overview of both technologies used in the context of pedestrian simulators can be found in Schneider and Bengler (2020). AV setups immerse the pedestrian in a virtual traffic scenario. However, their own, real body is visible in the VE. The pedestrian is surrounded by projection surfaces in a so-called CAVE (Cavallo et al., 2019). The size and number of displays varies greatly from setting to setting, as does the number of sides of the VE surrounding the participant. Ridene et al. (2015) used a floor and front projection to train children to safely cross the street. Meir et al. (2013) analyzed children's hazard perception when crossing the street using a 180° dome setup, while Cavallo et al. (2019) used a long corridor with projection screens on the left, front, and right sides to study crossing behavior in the elderly. Since none of these examples features a floor projection (except Ridene et al. (2015), however, here the sides are missing), these setups can rather be considered CAVE-like (Creagh, 2003; Schneider & Bengler, 2020). A CAVE-based pedestrian simulator can be found at the University of Leeds, which also features the largest walking space worldwide, with an area of 4×9 meters (Kaleefathullah et al., 2020).

In recent years, VR HMDs have been applied more frequently to study the behavior of pedestrians. The display is located directly in front of the participant's eyes. Thus, in contrast to CAVE-like systems, the user's own body is not visible, and the user is immersed in a completely virtual environment. This setup allows for a full field of regard, unlike in CAVEs, where for example a projection surface for the ceiling or rear may be missing. The amount of visual information that can be perceived at one time, however, is limited by the HMD's FoV. In the early 2000s, Simpson et al. (2003) introduced an HMD to study the street crossing behavior of younger pedestrians using VR. They used the Virtual Research Systems V8, featuring a 60° FoV and a resolution of 640×480 pixels per eye.

Since the introduction of a new generation of consumer-grade VR goggles, starting with the Oculus Developer Kit 1, the application of HMDs in pedestrian simulator studies has increased (Schneider & Bengler, 2020). These devices have sensors to determine the headset's position and rotation, but they rely on a PC to create the images (PC-based, see Table 1.1). The connection to the PC can be wireless, which increases the walking area. However, the space is ultimately limited by the tracking system. In addition to PC-based headsets, standalone systems were developed that integrate the computing unit into the headset and can thus be used independently. These systems are therefore more location-independent from cables or exterior tracking systems and allow for experiments outside the laboratory. Stadler et al. (2020) collected 463 individuals' preferences for designs of public transport waiting rooms by presenting them with a virtual prototyping environment using a standalone headset. The experiment was conducted in Singapore, Germany, and France. Mobile VR offers even more flexibility and cost efficiency, and thus scalability, when surveying large samples. A smartphone is inserted into a holder to display the VE (see Section 1.1). Schwebel et al. (2017) created a smartphone-based VE to teach children to safely cross the street on a large scale.

In addition to CAVEs and HMDs, monitor-based systems for displaying VEs should also be mentioned here. Depending on the number and size of the monitors and their arrangement, these systems can be compared with CAVEs. Depending on the setting, however, the immersion is limited. For example, Bart et al. (2008) used a single 17-inch monitor to train children to safely cross the street by displaying an avatar that was controlled via a keyboard.

With the advancement of technological possibilities, a variety of different settings for pedestrian simulators have been developed. However, the question arises regarding the extent to which the behavior observed in the simulators is equivalent to that in reality. Section 1.4 provides an overview of these and other validity issues, especially concerning pedestrian simulators.

1.4 Simulator Setups and Validity

Over the past century, the concept of validity has evolved in terms of complexity (Borsboom et al., 2004). A simple definition of validity is "whether a test really measures what it purports to measure" (Kelley, 1927, p. 14). The question then arises as to what factors influence whether a test fails to measure what it is supposed to. Campbell (1957, p. 297) further divides validity into whether "in fact the experimental stimulus makes some significant difference" (internal validity) and "to what populations, settings, and variables can this effect be generalized" (external validity). In other words, "internal validity is the extent to which we can draw confident causal conclusions, whereas external validity is the extent to which we can generalize the conclusions from our experiments to another context" (Lin et al., 2021, p. 2).

> “ When designing experiments and interpreting findings, the tension between experimentation goals and validity becomes apparent: Experiments provide the most direct way for determining causal effects and test theories because they maximize control and internal validity by simplifying, isolating, and making tractable even the most complex phenomena (Manzi, 2012; Pearl & Mackenzie, 2018), but these concessions are made at the cost of reducing external validity or the generalizability of the findings.
>
> (Lin et al., 2021, p. 2)

Simulators are often used as a safe and controllable alternative to real-world experiments (Deb et al., 2017; Feldstein et al., 2016; Mallaro et al., 2017; Schneider & Bengler, 2020), including when studying pedestrian behavior. The previous paragraph shows that a high degree of control is often accompanied by lower generalizability. In the context of pedestrian simulators, this means that simple traffic scenarios are often used to investigate effects in an isolated and controlled way, but the transferability of these results to real road traffic is limited (Schneider & Li, 2020). Additionally, artificial scenarios can limit realistic behavior. Slater (2009) differs in his taxonomy of presence between place illusion and plausibility illusion, that is, “the illusion that what is apparently happening is really happening (even though you know for sure that it is not)” (Slater, 2009, p. 3553). Slater (2009) argues that both place illusion and plausibility illusion are important to elicit realistic behavior.

In most pedestrian studies, a person’s behavior is examined in response to a certain stimulus, either in a simulator or in reality. If this stimulus produces a similar, equally directed effect in the simulation as well as in reality, it is called relative validity (Blaauw, 1982; Wynne et al., 2019). Furthermore, if the measured values in the simulator and reality match, one can speak of absolute validity (Blaauw, 1982; Wynne et al., 2019).

In addition to the design of the experiment, a number of technical factors also impact validity. Simulators usually try to represent reality in as detailed and as realistic manner as possible, but in the end they are only an approximation. The extent to which this approximation succeeds is mostly limited by technical factors (see Section 1.1). For example, the means of locomotion, a realistic perception of distance, or the representation of one's own body in VR can significantly influence behavior. These factors are examined individually next.

Locomotion in Virtual Reality

The most common use case for a pedestrian simulator is to determine the conditions under which a pedestrian would or would not cross a street, hence its commonly used synonym, "street-crossing simulator" (Bart et al., 2008; Cavallo et al., 2019). However, the technical composition of a pedestrian simulator limits the possibilities to cross a virtual street.

Early setups presented the VE on a single screen and featured keyboard input to navigate an avatar across the street (e.g., Bart et al., 2008). Keyboards and joysticks have been used as input devices in numerous studies (see Deb et al., 2017; Schneider & Bengler, 2020). Some studies used simple buttons to indicate the participant's intention to cross a street (e.g., Bernhard et al., 2008).

The emergence of immersive systems such as HMDs or CAVEs have also introduced new interaction modalities. The tracking of the head's position and rotation makes it possible to match the rendered perspective on the VE accordingly. Different technologies are employed that either use external sensors to detect the position of the head or headset (outside-in tracking) or are built directly into a headset and orient themselves independently in space (inside-out).

CAVEs mainly rely on an outside-in technology where the participant is equipped with markers, which may be mounted on a helmet, for example, and the head's rotation and translation are determined via a motion tracking system (e.g., Cavallo et al., 2019). Depending on their size, CAVEs either allow for naturalistic walking or the use of a treadmill, or do not allow for walking at all (Schneider & Bengler, 2020).

Earlier HMDs from the second wave of VR, such as the Oculus Rift and its predecessor DK2, functioned in a similar way, but with a mix of inside-out and outside-in tracking: External sensors tracked infrared lights mounted inside the HMD to track the position (outside-in). Rotation is measured with a built in (inside-out) IMU (Goradia et al., 2014). However, the maximum tracking space was limited to 2.4×2.4 meters (OculusVR, 2017), which does not allow for naturalistic walking, even in a single-lane street-crossing task. Researchers had to combine the HMD with an additional, professional motion tracking system to enable the physical crossing of the virtual street (Feldstein et al., 2016; Sween et al., 2017). In its second version, Valve's SteamVR tracking system supports tracking spaces of up to 10×10 meters (HTC Corporation, 2020), making additional motion tracking systems obsolete. The system works with photo diodes in the headset in combination with up to four infrared-emitting base stations, called "lighthouses," mounted on the walls (Holzwarth et al., 2021).

Low-cost, phone-based HMDs avoid the use of external sensors and instead rely on the IMU of the smartphone. However, this only allows rotation to be tracked. Currently, translational movements cannot be tracked reliably due to sensor drift. By contrast, mobile standalone HMDs additionally have cameras to compensate for the IMU drift and to detect translational movements with high accuracy (Holzwarth et al., 2021). Since no further external devices are necessary, natural locomotion is theoretically possible in an arbitrarily large area, limited only by the physical space in which one is located.

If the technology used or the available space does not allow crossing the road by walking, there are alternatives. For example, the intention alone to cross the road may be recorded by pressing a button (de Clercq et al., 2019). Likewise, techniques have been used that only mimic walking in place, such as swinging the arms (Orlosky et al., 2015).

Another possibility to increase the virtual explorable space within a limited physical area is to manipulate the user's perception, commonly referred to as redirected walking (RDW): "These methods create a distorted mapping of the VE by applying to the world subtle rigidbody and nonlinear transformations, respectively" (Sun et al., 2020, p. 286). For example, the ratio of physical to virtual rotation or translation can be manipulated, to generate so-called rotation or translation gains. Steinicke et al. (2008) report that translational gains of up or down to 22% (e.g., traveling 1.22 meters virtually while physically walking 1 meter) are still subtle enough to not be recognized.

Distance Perception in Virtual Reality

To transfer the results from a street-crossing simulator to reality, realistic distance perception is indispensable. As Section 1.2 has shown, the distance and speed estimation of relevant objects marks the beginning of every street-crossing decision. Three processes are responsible for distance perception, as explained by Armbrüster et al. (2008):

> "
> Pictorial, oculomotor, and binocular depth cues are combined to give an observer the three-dimensional impression of a scene. Pictorial depth cues are two dimensional, and the visual system interprets them in three-dimensional terms. Oculomotor depth cues comprise convergence and accommodation, which are dependent on each other and also on the binocular depth cue's disparity or stereopsis.
>
> (Armbrüster et al., 2008, pp. 9-10)

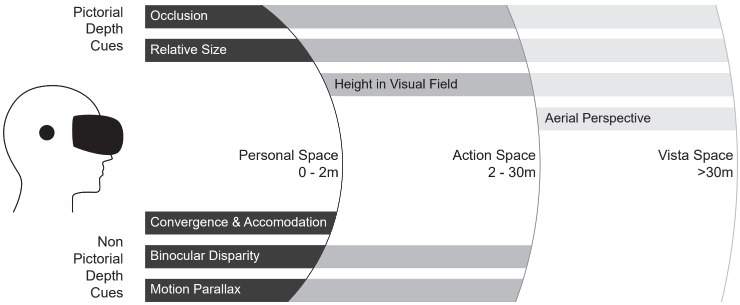

Figure 1.7: Influence of different depth cues on egocentric distance perception at various distances based on Renner et al. (2013).

Here, again, the influence of the technical composition becomes clear. If the technical setting of a simulator does not support stereo vision, depth perception is mainly achieved based on pictorial depth cues. However, this seems to be related to CAVE setups, since most HMDs feature either two separate displays or a single vertically split screen for stereo vision. The number of different available pictorial depth cues (e.g., shadows) as well as stereo vision seem to improve egocentric distance estimations (Hu et al., 2000), especially with nearby objects (Renner et al., 2013). Depending on the distance, the influence of the different depth cues varies (see Figure 1.7). If translational movements are possible (see Section 1.4), motion parallax can contribute even further, albeit mostly in terms of immersion and only little information to distance perception (Renner et al., 2013). Walking interaction in the VE can potentially improve distance estimates (Kelly et al., 2017). However, this positive effect depends on the method used to measure distance estimates (Kelly et al., 2017). Renner et al. (2013) provide an overview of applied measurement protocols and distance perception in VR. In general, egocentric distances are underestimated in VR (Renner et al., 2013). The underlying effects are varied and not yet fully understood (Renner et al., 2013). Newer HMDs seem to be less affected, although they continue to suffer from this distance compression effect (Kelly et al., 2017).

Embodiment in Virtual Reality

The role of presence and immersion have already been introduced in Section 1.1. With validity in mind, simulators often aim to be as immersive as possible to induce the feeling of being in the VE or the sense presence. One important aspect of presence is the display of the user's body.

While in CAVEs and AR setups, the participant can see their own body, this is not natively possible in VR with an HMD. One possible solution is to combine the HMD with a motion tracking system to display a virtual representation of the user's body, known as an avatar (Doric et al., 2016; Feldstein et al., 2016).

Steed et al. (2016, p. 1406) have found "a positive effect on self-report of presence and embodiment" when displaying an avatar, especially when the avatar is exposed to a threat. However, their study was conducted in a low-cost HMD setting without body tracking. Slater (2009) hypothesizes that a dynamic avatar (with body tracking) could contribute to both place illusion and plausibility illusion. The importance of an avatar for both concepts was later proven by Slater et al. (2010).

As mentioned above, simulators are a way of conducting experiments in a safe environment. However, they are also often a point of criticism: If passing vehicles are not perceived as a realistic threat, the observed behavior of the pedestrian in the simulation is called into question. The presentation of an avatar can potentially counteract this. The key factor is how the avatar is perceived. The avatar can encourage certain characteristics, such as immersion and presence, but the important thing is what feeling they evoke. De Vignemont

(2011) distinguishes between embodiment and the sense of embodiment: "Embodiment corresponds to a specific type of information processing, whereas the sense of embodiment corresponds to the associated phenomenology, which includes feelings of body ownership" (de Vignemont, 2011, p. 84). Based on the work of de Vignemont (2011) and Blanke and Metzinger (2009), Kilteni et al. (2012, p. 375) define the sense of embodiment "toward a body B [as] the sense that emerges when B's properties are processed as if they were the properties of one's own biological body."

The effect of avatars and embodiment is illustrated in the rubber hand illusion. In this experiment devised by Botvinick and Cohen (1998), the participant is seated with their left arm on a table. However, their view of the arm is blocked by a screen. Instead, the participant sees a rubber arm that is placed on the table. Both the artificial and the real hand are then stroked synchronously with a brush. Subjects report they increasingly perceive the rubber hand as their own. When the subjects are blindfolded and asked to indicate the position of the left hand, this location shifts increasingly toward the artificial hand (proprioceptive drift) depending on the duration of the stimulation. These effects disappear as soon as the haptic stimulus of the hands is applied asynchronously. This problem also occurred in Steed et al. (2016): In a low-cost setting without body tracking, participants saw an avatar tapping its hand to the music and were asked to do the same. However, this resulted in a lower sense of presence and body ownership, probably due to the asynchrony of the real and the avatar's tapping. Yuan and Steed (2010) recreated the rubber hand experiment and proved the illusion to work in VR. They also reported a strong physiological reaction when the virtual hand was exposed to a threat, but not when the virtual hand was replaced with an abstract arrow.

2

Scope of this Thesis

The introduction has shown that the technology in the field of VR and AR is subject to constant and ever more rapid change. The advancements have made it possible to use these technologies for the study of pedestrian behavior in the first place, but they also require continuous re-evaluation of the hardware currently in use. Only in this way can results from pedestrian simulators be interpreted and generalized to real-world traffic. In contrast to flight or driving simulators, which have been established, studied and, at least to some extent, validated in the past, the field of pedestrian simulators is still relatively young. In particular, the use of HMDs has gained momentum only with the commercialization of new VR hardware since 2016. However, this also means that many questions are still unanswered. The goal of this thesis is to answer some of these general questions in the specific context of pedestrian simulators. The focus of this work is not on exploring a single concept in detail; rather, the goal is to shed light on relevant questions from the research area of VR in the domain of pedestrian simulators and to set new impulses for the technical implementation. Which concepts should be considered to collect valid results and which innovations could enrich the field of pedestrian simulators in the future?

Section 1.4 has shown that fundamental concepts of a pedestrian simulator, such as locomotion in VR, distance perception, and embodiment, are highly susceptible to the changing technology. However, these effects cannot be considered in isolation; they are part of a complex interaction. This resulted in the motivation for Article 1: What influence do naturalistic walking and translation gains have on egocentric distance perception in a pedestrian simulator? What is the role of the applied method for measuring distance estimates (i.e., how can participants communicate their estimations) in the context of walking interactions and non-isometric mapping?

As described in Section 1.4, displaying an avatar in an HMD requires additional technical equipment. This additional effort often leads to the fact that an avatar is not displayed. However, technical innovations have reduced the effort considerably, and displaying an avatar might impact how one would react to the VE. Article 2 analyses the influence of an avatar with and without hand tracking on embodiment and aims to answer the question of whether the display of a virtual body representation impacts street-crossing decisions.

Article 3 deals with a technical evaluation of AR in the context of pedestrian simulators. Here, the question is, which new technical possibilities can meaningfully enrich the already established spectrum in the field of pedestrian simulators (see Section 1.3)? For this purpose, an AR pedestrian simulator is developed, evaluated and compared to street crossings in reality as well as against already established virtual settings.

In a similar fashion, Article 4 evaluates the suitability of a low-cost setup. Aside from technical influences, this work also aims to determine what influence the experimental environment and setting have on the test person. It answers the following question: Can experiments be conducted at home without a laboratory or an experimenter, and can this potentially even lead to more realistic behavior?

Article 1

3

Measuring Egocentric Distance Perception in Virtual Reality: Influence of Methodologies, Locomotion and Translation Gains

Summary Correct distance perception is essential to obtain credible results in pedestrian simulators. Participants often have the task of crossing a virtual street. It is crucial that the distance to the other side of the road and the time needed to cross the street are estimated realistically. This is complicated by an effect that often occurs in VR known as distance compression. Participants in previous studies have described this effect as "walking through honey." In an earlier study, the effects of locomotion and different translation gains on distance perception were investigated. Locomotion in this context means that the participants are able to walk the distance they later were asked to estimate. Translation gains manipulated the relationship between physically walked and virtually covered distance. However, the subjects only communicated their distance estimates verbally (Schneider et al., 2018).

In addition to locomotion and translation gains, this article investigates the effect of different measurement methods. Measurement methods describe different approaches to explicitly communicating the implicit distance perception. In addition to the verbal reports used in the preliminary study, other methods were identified in the literature, including the following: (1) visually guided walking, which is a variation of blind walking. In blind walking, participants are instructed to walk the estimated distance without any visual information. However, this can lead to participants walking at an angle instead of in a straight line. In the literature, this problem is addressed by, for example, providing acoustic or vibrotactile feedback as soon as the subject leaves a straight corridor. In the present work, all visual

stimuli were removed except for a horizon and an infinite straight line. (2) In another method, imagined timed walking, participants were asked to imagine walking and pressing a button while doing so. They were asked to release the button as soon as they had reached the target in their imagination. The duration of the button press was then multiplied by the individual walking speed to obtain a distance estimate. The individual walking speed was measured at the beginning of the experiment. (3) A third method is blind triangulated pointing, in which participants had the task of taking one step to the side and pointing to the imaginary target with a pointer. The estimated distance was then calculated using the angle of the pointer. Finally, there were two methods in which subjects communicated the estimated distance by throwing: (4) blind throwing, for which participants were given a 120-g sandbag to throw at the estimated target, and (5) virtual throwing as a variation of this: Instead of a physical object, a virtual ball was thrown via controller input.

The distances varied between 3 and 3.5 meters, which resulted in 2.4–4.2 meters for most extreme translation gains. Thirty participants (15 female, 15 male) experienced each of the five translation gains (0.8, 0.9, 1.0, 1.1, 1.2) and one non-walking trial block-wise for each measuring method, resulting in $(5+1) \times 6 = 36$ trials. This relates to five walking and one non-walking trial for each method. The study design and analysis plan were preregistered at Open Science Framework (https://osf.io/69skh/). Due to large variances in blind triangulated pointing, a separate analysis without this method was carried out. Contrary to expectations, lower misperceptions occurred with verbal estimates compared with the remaining methods. Differences between the non-verbal methods were also found. Measurements in which participants could walk to the target and back positively influenced walking-related measures (visually guided and imagined walking). Lower translation gains overall resulted in smaller estimation errors. The effect of translation gains was most prominent in walking-related methods.

Philipp Maruhn*	Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing
Sonja Schneider*	Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Writing – original draft, Writing – review & editing
Klaus Bengler	Funding acquisition, Supervision, Writing – review & editing

*Contributed equally

Maruhn, P., Schneider, S., & Bengler, K. (2019). Measuring egocentric distance perception in virtual reality: Influence of methodologies, locomotion and translation gains (H.-j. Sun, Ed.). *PLoS ONE*, *14*(10), e0224651. https://doi.org/10.1371/journal.pone.0224651

Article 2

4

Effects of Avatars on Street Crossing Tasks in Virtual Reality

Summary Pedestrian simulators can be classified into two categories with respect to their technical design, namely CAVEs and HMDs. In a CAVE, the test subject is located in a room where the walls consist of projection surfaces. A VE is then displayed on these projection surfaces. In most systems, the person's position is tracked. If the subject moves in the room, the display of the VE is adjusted accordingly. On the one hand, CAVEs are usually associated with high acquisition and maintenance costs because of their size and required hardware. On the other hand, relatively inexpensive HMDs allow subjects to experience the VE via lenses and a display directly in front of the eyes. However, this results in the subject no longer being able to see the physical environment, nor their own body. If a pedestrian simulator is supposed to reproduce realistic behavior, body sensation can play a decisive role.

The aim of this study is to investigate the effect of a virtual representation of one's own body (avatar) on crossing behavior in an HMD-based system. To represent an avatar, information regarding the orientation and location of the physical body parts is needed. In this work, an HTC Vive Pro is used as the HMD. Additionally, Vive trackers were attached to the feet and hips. Both hands were tracked via the controllers. Based on the tracked data, an inverse kinematic model was used to calculate the most probable position of the virtual extremities. These were subsequently displayed in the VE. Additionally, another avatar was created. In this instance, the handheld controllers were replaced by a "Leap Motion" controller attached to the HMD. This enables the virtual representation of finger movements and gestures in addition to the position of the two hands. 29 subjects (13 female,

16 male) experienced a virtual traffic scenario with each of the two avatars separately as well as in a baseline condition without an avatar. Virtual vehicles passed the subjects on a single-lane road at a constant speed of 30 km/h. During this process, vehicle gaps progressively increased from 1 to 6 seconds. Subjects were instructed to cross the road as soon as they judged the gaps to be sufficiently large and safe.

With an avatar, significantly smaller gaps were accepted than without one. No difference in gap acceptance was found between the two avatars. A learning effect could be demonstrated. In the second and third of the three experimental blocks, smaller gaps were accepted. However, this effect was counteracted with a randomized experimental design. A high number of virtual collisions occurred across all three conditions but without any significant difference between the avatar displays. In addition to the objective data, subjective perception was measured using presence and virtual body ownership questionnaires. The Leap Motion avatar produced a greater sense of virtual body ownership than the avatar without finger tracking. For the sensation of presence, no difference was found between the three conditions.

Philipp Maruhn Conceptualization, Data curation, Formal analysis, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing

Simon Hurst Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Validation — within the scope of a master thesis

Maruhn, P., & Hurst, S. (2022). Effects of avatars on street crossing tasks in virtual reality. In N. L. Black, W. P. Neumann, & I. Noy (Eds.), *Proceedings of the 21st congress of the international ergonomics association (IEA 2021)* (pp. 215–223). Springer International Publishing. https://doi.org/10.1007/978-3-030-74614-8_26

Article 3

5

Analyzing Pedestrian Behavior in Augmented Reality—Proof of Concept

Summary Road-crossing simulators have emerged as a tool to study pedestrian behavior under conditions that would not be possible in real-world settings. For many research questions, traffic observations, real-world experiments, or experiments on a test track are too complex, not controllable, or simply impossible. For example, encounters between pedestrians and autonomous vehicles are often investigated in virtual reality, since real prototypes do not yet exist. However, this requires that besides the actual object of interest (often only a single autonomous vehicle), the entire VE must be created as well. Participants thus often have high expectations regarding realism and the level of detail that must be fulfilled. The credo is that a high degree of immersion is necessary to be able to observe behavior as close to reality as possible. In this paper, an AR approach is presented.

The participants are situated on a test track and can see the real environment displayed via stereo video live-streamed in a head-mounted device (video pass-through). Virtual vehicles are then superimposed on this live stream. Thirteen subjects (6 female, 7 male) experienced in AR the same scenario as 30 other subjects with real vehicles: The subjects were standing at the curb of a 450-meter-long straight road. Two vehicles approached from the right with a constant speed of 30 or 50 km/h and a gap of 1–5 seconds. The subjects' task was to signal whether they judged the gap between the vehicles to be large enough to cross the road safely. This signal was given by stepping forward without actually crossing the street. Subjects were asked to do this as

soon as they would initiate the crossing after the first car had passed. Each participant experienced each combination of the two levels of speed and five gap sizes twice, resulting in 20 trials per participant. The trial order was randomized.

A logistic mixed-effects analysis revealed that in the AR condition, significantly larger and thus overall fewer gaps were rated as being safe to cross. The modeling of the acceptance rates shows that in both conditions, namely the real world and AR, a similar relationship exists between gap size and acceptance, but these are offset. The time at which the subjects initiate the crossing is also delayed in AR. Finally, the test persons were asked about their subjective impression of the street-crossing task by means of a questionnaire. Here, no differences were found in AR compared with the experiment conducted in reality.

Since this work is a first proof of concept, further potential for improvement was identified. For example, information on geolocation, time of day, and current weather conditions could be used to calculate a realistic shadow cast by the virtual vehicles. This paper also discusses different approaches to match the position of the vehicles with the real environment. With more accurate calibration methods, the vehicles could be even more realistically mapped on the road, especially at greater distances. However, it must be critically questioned whether this is necessary. The limited resolution of the HMD, particularly in a video pass-through, reduces the virtual vehicles at large distances to a few poorly visible pixels. Overall, it could be shown that AR is a promising tool to investigate pedestrian behavior. Future studies can build upon the presented work and identify suitable use cases and research questions.

Philipp Maruhn	Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Writing – original draft, Writing – review & editing
André Dietrich	Writing – review & editing
Lorenz Prasch	Visualization, Writing – review & editing
Sonja Schneider	Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Writing – review & editing

Maruhn, P., Dietrich, A., Prasch, L., & Schneider, S. (2020). Analyzing pedestrian behavior in augmented reality — Proof of concept. *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 313–321. https://doi.org/10.1109/VR46266.2020.00051

Article 4

6

VR Pedestrian Simulator Studies at Home: Comparing Google Cardboards to Simulators in the Lab and Reality

Summary CAVEs and HMDs are the two most commonly used technologies for pedestrian simulators. Even though consumer-grade hardware is mostly used as HMDs, high-end products are usually employed in this context as well. Although the costs are significantly lower compared with CAVEs, they are also not negligible due to the additional required hardware (e.g., high performance PCs). Most pedestrian simulator studies are performed in laboratories at universities or research institutes, and test persons are consequently recruited from the immediate environment, at least as long as no specific demography is part of the research question. The aim of this paper is to evaluate cost-effective alternatives. In addition, it examines whether pedestrian simulator studies are feasible in a remote setting to provide VR studies with access to a broader subject population.

For this purpose, the gap acceptance study of Article 3 was replicated as a smart phone application. Again, subjects had the task of rating gaps between two vehicles regarding their willingness to cross the street. In a between-subjects design, 30 subjects (15 male, 15 female) in each condition performed the experiment either in a laboratory, at the university, or at home. A supervisor was present in the laboratory, but not at home. In both conditions, the app was sufficiently self-explanatory to be able to perform the experiment without assistance. After an introduction within the application, the subjects were asked to place their smartphone in the provided Google

Cardboard device. With its help, the VE was displayed. The data was then stored in an online database and compared with the results of previous studies conducted in reality on the test track, in a CAVE, using an HMD (HTC Vive Pro), and in AR (see Article 3).

The technical peculiarities of the cardboard device necessarily led to some modifications in the experimental design. Instead of a step toward the street, the participants signaled a crossing initiation by pressing a button. Due to a high level of discomfort revealed in pilot studies when the cardboard was worn, the number of trials was halved. Each combination of speed (30 or 50 km/h) and gap size (1–5 seconds) was displayed once (cf. Article 3) in a randomized order.

The data was analyzed using a Bayesian approach, and the results paint a similar picture for the two cardboard conditions when compared with the other simulators. Fewer gaps were accepted than in reality, with acceptance rates ranging between CAVE and HMD. As in the other simulators, subjects made the decision to cross the road later than in reality. Neither differences nor equivalence could be demonstrated between the two cardboard settings in the laboratory and at home. Following the experimental trials, subjective assessments were again collected by means of a questionnaire in the app. Here, it must be noted that the subjects in the remote condition indicated a higher agreement with their everyday behavior for trials in which they decided not to cross the street. For trials in which they would have crossed the road, a similar trend was observed, but no difference could be verified. Overall, it appears that cardboard devices are a suitable, cost-effective, and therefore scalable alternative to high-end HMDs. However, technical limitations in terms of display quality and interaction possibilities restrict the range of applications.

Philipp Maruhn	Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing

Maruhn, P. (2021). VR pedestrian simulator studies at home: Comparing google cardboards to simulators in the lab and reality. *Frontiers in Virtual Reality*, *2*. https://doi.org/10.3389/frvir.2021.746971

7

Discussion: Revisiting Street Crossing Simulators

VR and AR are being applied in a wide variety of areas, for example in the automotive industry and healthcare as well as in culture and the entertainment market (Bitkom e.V., 2021). In the past, these technologies were reserved for specific research disciplines. Today, the high availability of new and inexpensive devices has led to their widespread use in science as well. As described in Section 1.2, research into pedestrian behavior has great potential to increase traffic safety through the use of VR and AR. With the emergence of new consumer-grade HMDs in the second wave of VR, these HMDs have also been increasingly used in pedestrian simulators alongside the already established CAVEs. This has allowed a growing number of researchers to study pedestrian behavior in safe and controlled environments. However, despite the wider availability and subsequent application of these new tools, a large number of questions remain unanswered.

This work transfers relevant research questions from the broad field of VR into the context of pedestrian simulators while discussing the appropriateness of new technologies and methods. In particular, the area of HMDs, whether for VR or AR, is characterized by fast-moving technological advancements (Dörner et al., 2019). This development is of tremendous potential for the use of HMDs in behavioral research. In the past four years of this work alone, the following has been observed:

- The emergence of new tracking technologies, making the use of expensive and complex additional motion-tracking systems obsolete;

- The continuous expansion of tracking space, allowing even for naturalistic street crossings with consumer-grade systems;
- A dramatic increase in FoV, display resolution, and image quality;
- Eye tracking finding its way into commercial systems.

These rapid technological advances highlight that this work can only be a snapshot of the current state of the art. However, this does not diminish the importance of this work, but rather highlights the need for continuous re-evaluation. In the following discussion, the limitations of current settings, their implications for the interpretation of results from pedestrian simulators, and the resulting research questions and needs for action are presented. Subsequently, the use of specific hardware appropriate to different research questions is discussed.

7.1 Current Limitations

Section 1.2 shows that a complex perceptual and information-processing procedure (see Figure 1.6) underlies each road crossing. To translate observations from the simulator to reality, it must be understood whether these processes occur in the same way in VR or what differences may exist.

Article 1 reconfirms the findings in the literature (Renner et al., 2013), albeit to a lesser extent, that distances are underestimated in VR. This effect was also demonstrated in an AR video pass-through HMD (Pfeil et al., 2021). Among other things, a restricted FoV may promote this effect (Kelly et al., 2017; Pfeil et al., 2021; Willemsen et al., 2009), but improved display weight and resolution also seem to improve distance perception (Kelly et al., 2017; Willemsen et al., 2009). This is one explanation for why this issue is less of a problem in newer HMDs.

Kelly et al. (2018) further report that walking interaction can positively influence distance perception. This is directly related to the question of how users can communicate their crossing decision in a pedestrian simulator. In the past, different techniques have been used (see Section 1.4). Te Velde et al. (2005) report that verbal judgment tasks differ from actual street crossings. Overall, it should be noted that naturalistic walking appears to be the most suitable for a road crossing in the simulation and that the current state of technology allows for that to some extent. However, this is only true for a limited space with spatially small scenarios.

Translation gains can artificially increase the virtual space and potentially reduce distance compression (see Article 1), albeit only in a limited range. The research interest in solving the problem of locomotion in an endless virtual space bounded by a physical space is enormous, and a variety of approaches exist (see Nilsson, Peck, et al., 2018; Nilsson, Serafin, et al., 2018). In the future, it will be necessary to evaluate whether additional technologies such as treadmills or redirected walking are suitable for pedestrian simulators in more complex scenarios. Redirected walking refers to different methods of creating the illusion of walking along a straight line in VR when, for example, in fact one is actually walking along the path of an arc. However, such techniques must be used with caution in the context of pedestrian simulators for two reasons. First, they pose the danger of additional interference with the naturalness of the crossing movement, which also complicates the interpretation of absolute metrics (e.g., walking speed and safety measures such as post encroachment time). Second, they could cause additional problems such as simulator sickness.

Simulator sickness, or cybersickness (CS), is a physical response (similar to motion sickness) that may occur during or after the use of VR applications (Biocca, 1992; Dużmańska et al., 2018). In contrast to motion sickness, CS can be triggered even without physical movement, for example by simulated movements (i.e., vection) in the VE (Biocca, 1992) or by time delays between head movement and rendering in VR

(Dużmańska et al., 2018). In general, HMDs seem to cause CS more often than CAVEs (Pala et al., 2021b; Weidner et al., 2017), although some studies have found no difference between the two systems in terms of CS (Borrego et al., 2016; Pala et al., 2021a). The continuous improvement of tracking technology and display quality in HMDs can potentially reduce CS. However, CS remains a problem (Dużmańska et al., 2018).

Given that negative symptoms increase with VR exposure time (Deb et al., 2017; Dużmańska et al., 2018), experiments should be time-limited. However, a trade-off must be made, since repetitive tasks in pedestrian simulators are subject to a learning effect (see Article 2) and, therefore, require a sufficiently long familiarization phase. Furthermore, the goal should always be to reduce CS as much as possible. In addition to ethical reasons, the inverse relationship between CS and presence should be considered (Weech et al., 2019):

> “When taken together, the evidence [...] begins to clarify the type of relationship that exists between presence and CS:
>
> – Approaches that reduce sensory mismatch show potential for reducing CS and increasing presence;
>
> – Both presence and CS are increased by the addition of stereoscopy, high field-of-view display conditions, and by enhancing the likelihood that a display will evoke vection;
>
> – Increasing factors such as intuitiveness of interaction and control of navigation lead to higher presence and lower CS;
>
> – Men and individuals with more gaming experience demonstrate lower CS and higher presence, although the partial effects of sex and gaming are not fully clear.
>
> (Weech et al., 2019, p. 13)

Presence is of high value in the VR domain in general. It is also of great importance in the field of pedestrian simulators if one considers how often questionnaires about presence are applied in this context (Bhagavathula et al., 2018; Deb et al., 2017; Feldstein et al., 2016; Feng, 2021; Pala et al., 2021b; Ye et al., 2020). Extensive research has been conducted on presence, how to increase it, and how to measure it. However, there is still a lack of knowledge about how presence affects human responses in VEs, such as when VR is applied as a therapeutic tool in psychology (Schuemie et al., 2001). The same is true for pedestrian simulators. Article 2 shows that displaying an avatar has no significant effect on self-reported presence in the given scenario. Nevertheless, it affects street-crossing behavior, with smaller gaps being taken when an avatar is visible. The question is whether high presence scores should be considered the gold standard, or if individual aspects and their effect on behavior in a pedestrian simulator should be investigated instead.

Many studies describe roughly how their VE is composed and also which display device is used. However, little knowledge is available about which aspects are relevant, and no standards exist regarding which features must be included as a minimum. Some aspects will be solved by technological progress. Increasingly, the FoV and display quality of many devices are approaching the human limits of perception. However, some questions remain unanswered: How realistic should the sound and trajectory of passing vehicles be? What role does haptic feedback play? For example, can standing on a real curb (as in the virtual pit experiment by Meehan et al., 2002) induce more realistic behavior?

What is the role of on-boarding in a virtual street-crossing experiment? In the case of a driving simulator, a mental preparation for the task at hand can be assumed by physically entering the mock-up, but VR studies hardly use such real-world metaphors. Deb et al. (2017) use a familiarization scene consisting of a small room with a desk and blackboard before the participants experience the city scenario in their pedestrian simulator. Methodological standards must also be defined, for example how to deal with spectacle wearers. Consequently, the display distance (eye relief) must be set the same for all subjects. This reduces the FoV for larger distances or excludes spectacle wearers as participants for smaller distances. This aspect is usually not described in the study designs.

7.2 What Hardware to Use

The newest! At least, that is what one would assume from most of the content described here. The key phrase "technological advancements" and the positive influence of new technology on immersion and presence has been mentioned on countless occasions in this thesis. A similar picture emerges by reviewing the literature. For this purpose, studies with the keywords "pedestrian" and "HMD" or "head-mounted display" were collected from SCOPUS and Google Scholar dated from 2013 onwards (i.e., the release year of the Oculus Rift DK1 and the beginning of a new generation of HMDs).

Figure 7.1 compares the year of publication and the age of the HMD used. It becomes apparent that the old generation of HMDs has eventually been superseded by new and commercial products. When interpreting the data, it should be noted that there can sometimes be a long period of time between the collection of the data and the publication of the study. Most of the studies do not specify the year of data collection. For the few known cases, however, this delay can be up to four years.

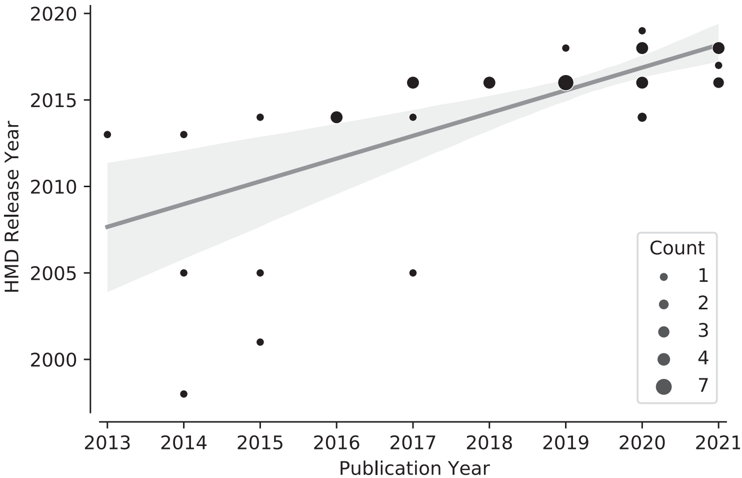

Figure 7.1: Publication year of studies since 2013, which analyzed pedestrian behavior using HMDs compared to the release year of the HMD in use. The line represents the regression model of HMD release year as a function of the publication year, the grey area the 95% confidence interval. Sources are below.

	HMD	HMD Release
Nuñez Velasco et al. (2021)	HTC Vive	2016
Feng (2021)	HTC Vive	2016
Pala et al. (2021a)	HTC Vive Pro	2018
Pala et al. (2021b)	HTC Vive Pro	2018
Schneider et al. (2021)	HTC Vive Pro	2018
Camara et al. (2021)	HTC Vive Pro	2018
Epke et al. (2021)	Oculus Rift CV1	2016
Wirth and Warren (2021)	Samsung Odyssey	2017
Lubetzky et al. (2020)	HTC Vive	2016
Deb et al. (2020)	HTC Vive	2016
Maruhn et al. (2020)	HTC Vive Pro	2018
Fuest et al. (2020)	HTC Vive Pro	2018
Vizzari (2020)	Oculus Go	2018

Tram and Parker (2020)	Oculus Quest	2019
Kobayashi and Yoshioka (2020)	Oculus Rift CV1	2016
Schneider and Li (2020)	HTC Vive Pro	2018
Luo et al. (2020)	Oculus Rift CV1	2016
Feldstein and Dyszak (2020)	Oculus Rift DK2	2014
Azam et al. (2020)	Oculus Rift DK2	2014
Stadler et al. (2019)	HTC Vive	2016
Lee et al. (2019)	HTC Vive	2016
Schneider et al. (2019)	HTC Vive Pro	2018
Hudson et al. (2019)	HTC Vive	2016
Savino et al. (2019)	HTC Vive	2016
Perez et al. (2019)	Microsoft HoloLens	2016
Kooijman et al. (2019)	Oculus Rift CV1	2016
de Clercq et al. (2019)	Oculus Rift CV1	2016
Bhagavathula et al. (2018)	HTC Vive	2016
Dietrich et al. (2018)	HTC Vive	2016
Otherson et al. (2018)	HTC Vive	2016
Deb et al. (2018)	HTC Vive	2016
Mallaro et al. (2017)	HTC Vive	2016
Deb et al. (2017)	HTC Vive	2016
Böckle et al. (2017)	HTC Vive	2016
Chang et al. (2017)	HTC Vive	2016
Sween et al. (2017)	Oculus Rift DK2	2014
Corbett and Morrongiello (2017)	VR1280	2005
Ishii et al. (2016)	Oculus Rift DK2	2014
Rojas et al. (2016)	Oculus Rift DK2	2014
Feldstein et al. (2016)	Oculus Rift DK2	2014
Doric et al. (2016)	Oculus Rift DK2	2014
Orlosky et al. (2015)	Oculus Rift DK2	2014
Morrongiello and Corbett (2015)	VR1280	2005
Ragan et al. (2015)	nVis SX111	2001
Rojas et al. (2014)	Oculus Rift DK1	2013
Morrongiello et al. (2014)	VR1280	2005
Vilar et al. (2014)	Sony HMD	1998
Rojas and Yang (2013)	Oculus Rift DK1	2013

Overall, the general impression is confirmed that the currently best-performing hardware available is used at the time of data collection. In more recent studies in particular, “best performing” often means newest hardware. In the previous chapter, however, we noticed that this might not be the decisive criterion at all.

Lessons Learned

Article 3 and Article 4 compare the technologies used for different simulator concepts. The experiments are based on a relatively simple use case. The subjects indicate whether the gap between two vehicles seems large enough to cross the road safely. Since the test was also conducted in reality, it was not possible to perform a real crossing, even though it would otherwise be desirable to have participants actually walk across the street (see te Velde et al. (2005), Section 1.4). It can be concluded that for this use case, a low-cost, low-tech solution achieves similar results to current state-of-the-art simulators. Neither objective values such as the number of accepted gaps and crossing initiation time, nor subjective experience (e.g., how likely a collision would have been) show large differences between the simulators. However, it should be noted that all simulators deviate from reality, especially in the objective metrics. This indicates that the latest hardware might not always be necessary. However, as mentioned, these findings are limited to a very specific use case in a simple street-crossing scenario.

In most pedestrian simulator studies, not only is the scenario simple (Schneider, 2021), but it also often only involves a single user. However, research questions also arise in the context of pedestrian simulators that require interactions of multiple users in a shared VE. This can be illustrated by looking at the research of interactions between pedestrians and autonomous vehicles in VR. There has been a drastic increase in the number of VR studies on how pedestrians perceive autonomous vehicles (Burns et al., 2020) and how driverless vehicles could communicate with pedestrians via new external displays (Deb et al., 2018; Dietrich et al., 2018; Otherson et al., 2018; Stadler et al., 2019). Probably one of the reasons for the increase was that the new generation of HMDs made it much easier to get started with VR, both in terms of monetary costs and regarding the ease of creating the VE using game engines such as Unity and Unreal. However, most of these studies have several shortcomings. Aside from insufficient traffic complexity (see Section 1.4), most studies are limited in terms of how an autonomous vehicle can communicate with a single pedestrian (Colley et al., 2020). Article 2 has shown that with little effort an avatar can be displayed in VR. This would potentially allow researchers to investigate issues with multiple pedestrians as well. Another possibility would be the use of non-player characters (NPCs). However, their behavior must be carefully designed to evoke natural reactions.

Article 2 has also shown that hand and finger tracking can improve the illusion of virtual body ownership. However, this has not led to any significant changes in street-crossing behavior, presumably because one's own hands are only visible in the peripheral FoV when walking, if at all, and probably not at all due to the limited FoV in HMDs. Except for some specific research questions (such as when pedestrians interact with vehicles using hand gestures), a simple avatar without finger tracking seems to be sufficient.

Article 3 describes how an AR pedestrian simulator can be created with little effort. However, it also shows that spatially constrained scenarios are more suitable and can avoid problems with calibration and poor stereo camera quality. The main advantages when compared with mere VR are that the participant sees their own body and no VE needs to be created except for the subject of the study itself. Especially for those research questions regarding the field of interaction of pedestrians with autonomous vehicles, an AR approach could be suitable.

Another intriguing point is highlighted in Article 4: the scalability of studies. In a review of pedestrian simulator studies, Schneider and Bengler (2020) found an average sample size of 69.5 participants. However, this number is heavily skewed by a small number of studies with large sample sizes and studies with between-subject designs (Schneider & Bengler, 2020). A large proportion of studies are characterized by small sample sizes.

Scalable hardware and new recruitment methods might solve this issue. It has been shown that for a simple application, even a low-cost solution can be suitable. In addition, experiments can be conducted outside the laboratory. Stadler et al. (2020) carried out a large sample experiment at public events across three countries with a standalone HMD. Steed et al. (2016) performed an experiment "in the wild" to measure presence with and without an avatar in low-cost HMDs. They invited HMD owners via websites and email to download the application and to participate in their study. In the end, Steed et al. (2016) collected 59 complete data sets, which correspond to a return rate of 15%. In line with Steed et al. (2016), Article 4 confirmed the feasibility of such remote studies. However, the two studies relied on different approaches to recruit participants. Steed et al. (2016) presupposed that the participants have the necessary hardware at hand.

Although the number of private users continues to grow, if one considers their demographics, recruiting this way could lead to a very specific sample profile (Kelly et al., 2021). Shipping Google Cardboards as described in Article 4 is one solution to reach participants outside this user group. However, it has also been shown that under certain circumstances, either low response rates are to be expected (Steed et al., 2016) or the subjects need a reminder to perform the experiment (Article 4).

Table 7.2: Suitability criteria of AR and VR setups for application as pedestrian simulators.

Criterion	Explanation
Ergonomics	Can the hardware be adapted to the user?
Cybersickness	Is the system usable without causing symptoms of cybersickness?
Avatar	Is the own body visible or virtually depictable?
Scalability	Is the system suitable for large-scale decentralized studies?
Locomotion	Is large-scale naturalistic walking possible?
Multi-user	Can multiple users simultaneously experience the same VE?

Identifying Suitable Hardware

In summary, it can be stated that different hardware is differently suited for different research questions. Figure 7.2 shows the strengths and weaknesses of different pedestrian simulator setups in relation to the criteria listed in Table 7.2.

In the following explanation, the proficiencies in the individual dimensions are justified. In this context, the individual scales are not to be understood as linear, but as ordinally scaled differences between the systems.

Ergonomics As the name suggests, HMDs are attached to the user's head. However, since the anthropometric characteristics of users can vary greatly from one individual to another, a number of ergonomic factors must be considered. Especially for special user groups such as children or the elderly, ergonomic criteria can be a decisive factor. Heavier HMDs can cause discomfort in the area of the neck or when

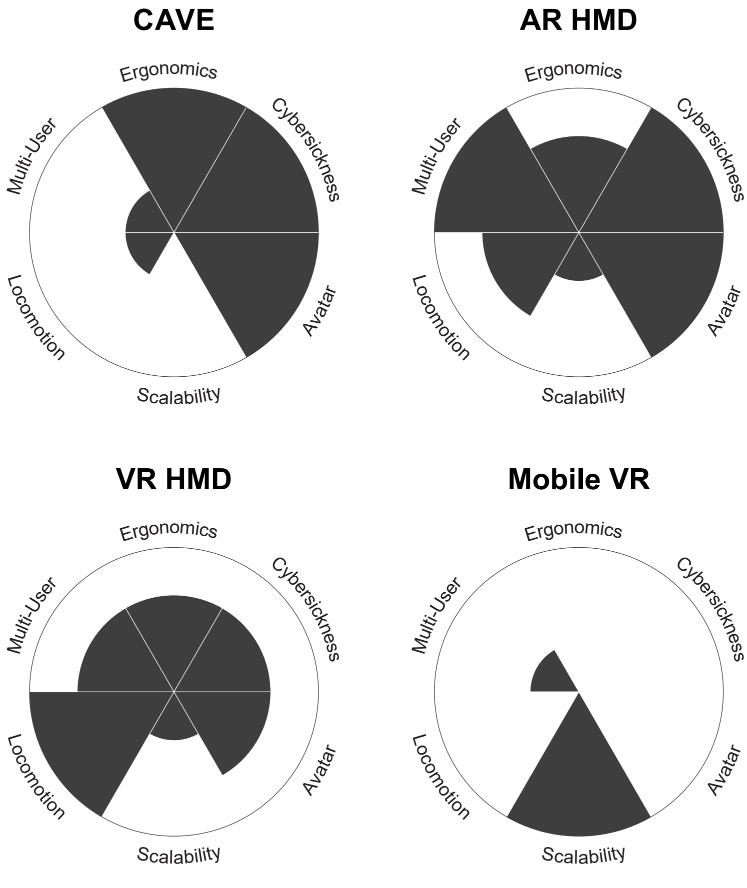

Figure 7.2: Suitability of different pedestrian simulator setups within distinctive feature characteristics (see Table 7.2). The size of the sectors denotes the performance and suitability in the respective category.

resting on the user's nose. Many devices are not designed for the small heads of children. On the one hand, this can cause problems when attaching them to the head, and on the other hand, the adjustment range of the lenses is usually not sufficient to accommodate for the smaller interpupillary distance of children (Geruschat et al., 1999; Pala et al., 2021b). The same restrictions apply to (video pass-through) AR HMDs. Usually, the customization options of mobile VR HMDs are additionally limited. Lower-priced models (e.g., Google Cardboard) in particular often lack any form of ergonomic adjustment. The absence of cushioning can cause additional discomfort. These restrictions may limit the reasonable duration of experiments (see Article 4). CAVEs, by contrast, are not subject to any of the above-mentioned restrictions.

Cybersickness In general, there is a risk of CS with all the systems described here. However, as for the other criteria, relative differences can serve as a basis for deciding which system to choose. The prevalence of CS depends not only on the user's individual factors and the software (composition and interaction with the VE) but also on the hardware in use (Rebenitsch & Owen, 2020). Aside from "nausea" and "oculomotor", "disorientation" is one of the three symptom clusters related to CS (Davis et al., 2014). "Disorientation" poses a greater risk with systems that completely block the real-world view (VR HMD) compared with settings in which the real world (AR) or at least parts of it (e.g., one's own body or missing projection surfaces in CAVEs) are still visible. Especially for the elderly, "who are typically thought of as being more sensitive to CS" (Pala et al., 2021b, p. 9), disorientation might be a contra-indicator. However, a direct comparison of a CAVE-based simulator vs. an HMD-based pedestrian simulator did not show a significant difference on CS between either system, neither for the elderly nor for younger adults or children (Pala et al., 2021a, 2021b). The studies did, however, reveal a higher susceptibility to CS for children compared with young adults, independent of the simulator device (Pala et al., 2021a).

Phone-based systems (mobile VR) cannot compete with commercial VR HMDs in terms of display quality or tracking accuracy. As a result, oculomotor-related symptoms such as eye strain might increase. Additionally, tracking errors could result in vection, as evidently reported in Article 4: A sensor drift in the IMU created a virtual head rotation without physically turning the head, eventually resulting in increased CS.

Avatar Depending on the research question, the presentation of an avatar may or may not be mandatory, but in most cases it is at least advisable (see Article 2). In CAVEs and in AR, one's own body can be seen, which corresponds to the preferred situation. In VR, a virtual body replica can be displayed with the help of additional sensors. In mobile VR, this is only possible, if at all, to a very limited extent.

Scalability As simplicity increases, so does the cost-effectiveness of setups and thus scalability, enabling decentralized studies (see Article 4). Mobile VR in particular stands out here because of its low cost. By contrast, CAVEs are completely location-bound. HMD-based systems are, in general, location-independent. However, their high acquisition costs reduce their scalability compared with, for example, Google Cardboards. Scalability is usually negatively correlated to the interaction possibilities: for instance, whether a street crossing can be performed by means of natural walking movements.

Locomotion Natural walking is usually advisable to make metrics such as post encroachment time and virtual collisions (see Article 2) measurable in the first place. In addition, natural crossing seems to have a positive effect on distance perception (see Article 1). In mobile VR, however, virtual movement is currently only possible indirectly via controller inputs. Natural walking is always limited by the physical space available. CAVEs are especially affected by this because of the limitation of the projection screens. Standalone HMDs have a theoretically unlimited tracking area. The latter also allow for more complex scenarios than a simple road crossing. AR settings are ultimately space-bound by the real world scenario in which they take place.

Multi-user As described earlier, it can be assumed that scenarios with multiple pedestrians will be increasingly investigated by researchers. In general, CAVEs are collaborative workspaces (Mestre, 2017). In the context of pedestrian simulators, however, they are only suitable for multi-user studies to a limited extent, since the currently displayed perspective is usually only synchronized to a single user (Mestre, 2017). In VR, a multi-user environment could be realized by displaying avatars. As displaying an avatar is very limited in mobile VR, so is a multi-user setting. AR offers the advantage that other users can be seen in real life.

In addition to the dimensions described above, there are further decision criteria. A key drawback of the AR setup presented in Article 3 is that it requires a real-world environment. Although this is associated with a low effort compared with the creation of the VE, and a high degree of realism, the variability of scenarios is considerably limited. In the future, a mixture of approaches would be conceivable, for example AR in green boxes to represent one's own body in the VE by means of chroma keying. Depending on the research question, further individual aspects must also be discussed. Additionally, the weighting of the individual criteria should be adjusted from case to case. Nevertheless, Figure 7.2 provides a basis for discussing and evaluating the different hardware aspects of pedestrian simulators.

Finally, regarding the interpretation and generalization of pedestrian simulator study findings, overall, the impression is that effects in reality can also be found in VR (relative validity). However, Article 3 and Article 4 have confirmed that there is no absolute validity for any of the evaluated systems. This is in line with findings by Schneider (2021). Similar findings can be found in Fuest (2021). Here, a Wizard of Oz video and VR approach were compared to evaluate the trajectories of autonomous vehicles from the perspective of pedestrians. Again, the same effects were found in VR and reality (relative validity), but the absolute measured values cannot be transferred directly (no absolute validity).

8

What the Future Holds—Toward the Ultimate Pedestrian Simulator

VR and AR have come a long way since Sutherland (1965) envisioned the "ultimate display." Both technologies have now established themselves in science as research tools and are becoming increasingly widespread in the consumer sector. In the field of pedestrian research, VR has become an indispensable tool. In this context, applied science is primarily a beneficiary of the ongoing development of commercial headsets. Therefore, it is natural to ask where this development is heading and how it will benefit research into pedestrian behavior. A number of trends can be discerned. Display quality will continue to increase; the first industrial grade manufacturers are already claiming to produce displays with human-eye resolution (Varjo, 2021). In the consumer market in particular, an increase of standalone headsets can be observed, which could lead to an even easier adoption of VR in other fields. This could potentially enable an unlimited tracking range and large-scale, decentralized behavioral studies.

These technological improvements will be further driven by computer vision algorithms, which are increasingly enhancing camera-based tracking. In combination with Fifth Generation Cellular Network Technology (5G) technology, these advances can potentially push forward phone-based AR. With both better sensors and algorithms, the problem of occlusion in AR can also potentially be solved more easily, turning AR into a valuable tool in pedestrian research and in ergonomics overall.

Pedestrian simulators are an indispensable tool on the road to zero traffic fatalities. On the path to the ultimate pedestrian simulator, however, hardware issues are not the only obstacles to be overcome. The simulated scenarios themselves must approximate real road traffic more closely to be able to transfer findings from the simulation to everyday road traffic with as little bias as possible. However, to do so, the problem of infinite locomotion must be solved. Moreover, to make studies more comparable, the methodological procedures should be standardized.

In an ultimate pedestrian simulator, the behavior of pedestrians in the simulation would be indistinguishable from reality. Nevertheless, the environment is controllable enough to investigate specific hypotheses and scenarios. The quest for the ultimate display requires narrowing this gap between the world as it could be and the world as it is to ultimately pave the way for a world as we imagine.

Bibliography

Altick, R. (1978). *The shows of london*. Belknap Press.

Armbrüster, C., Wolter, M., Kuhlen, T., Spijkers, W., & Fimm, B. (2008). Depth perception in virtual reality: Distance estimations in peri- and extrapersonal space. *CyberPsychology & Behavior*, *11*(1), 9–15. https://doi.org/10.1089/cpb.2007.9935

Avila, L., & Bailey, M. (2016). Augment your reality. *IEEE Computer Graphics and Applications*, *36*(1), 6–7. https://doi.org/10.1109/MCG.2016.17

Avineri, E., Shinar, D., & Susilo, Y. O. (2012). Pedestrians' behaviour in cross walks: The effects of fear of falling and age [Safety and Mobility of Vulnerable Road Users: Pedestrians, Bicyclists, and Motorcyclists]. *Accident Analysis & Prevention*, *44*(1), 30–34. https://doi.org/https://doi.org/10.1016/j.aap.2010.11.028

Azam, M., Choi, G., & Chung, H. C. (2020). An HMD-based pedestrian simulator for training and measuring individuals perceptual-motor behaviour in road crossing. *International Journal of Human Factors and Ergonomics*, *7*(4), 325. https://doi.org/10.1504/IJHFE.2020.112501

Azuma, R. T. (1997). A survey of augmented reality. *Presence: Teleoperators and Virtual Environments*, *6*(4), 355–385. https://doi.org/10.1162/pres.1997.6.4.355

Bart, O., Katz, N., Weiss, P. L., & Josman, N. (2008). Street crossing by typically developed children in real and virtual environments. *OTJR: Occupation, Participation and Health*, *28*(2), 89–96. https://doi.org/10.3928/15394492-20080301-01

Barton, B. K., & Schwebel, D. C. (2007). The roles of age, gender, inhibitory control, and parental supervision in children's pedestrian safety. *Journal of Pediatric Psychology*, *32*(5), 517–526. https://doi.org/10.1093/jpepsy/jsm014

Bengler, K., Drüke, J., Hoffmann, S., Manstetten, D., & Neukum, A. (2018). *Ur:ban human factors in traffic*. Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-15418-9

Bernhard, M., Grosse, K., & Wimmer, M. (2008). Bimodal task-facilitation in a virtual traffic scenario through spatialized sound rendering. *ACM Trans. Appl. Percept.*, *8*(4). https://doi.org/10.1145/2043603.2043606

Bhagavathula, R., Williams, B., Owens, J., & Gibbons, R. (2018). The reality of virtual reality: A comparison of pedestrian behavior in real and virtual environments. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *62*(1), 2056–2060. https://doi.org/10.1177/1541931218621464

Biocca, F. (1992). Will simulation sickness slow down the diffusion of virtual environment technology? *Presence: Teleoperators and Virtual Environments*, *1*(3), 334–343. https://doi.org/10.1162/pres.1992.1.3.334

Bitkom e.V. (2021). Augmented und virtual reality: Potenziale und praktische anwendung immersiver technologien (Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e. V, Ed.). Retrieved September 1, 2021, from https://www.bitkom.org/Bitkom/Publikationen/Augmented-und-Virtual-Reality

Blaauw, G. J. (1982). Driving experience and task demands in simulator and instrumented car: A validation study. *Human Factors*, *24*(4), 473–486. https://doi.org/10.1177/001872088202400408

Blanke, O., & Metzinger, T. (2009). Full-body illusions and minimal phenomenal selfhood. *Trends in Cognitive Sciences*, *13*(1), 7–13. https://doi.org/https://doi.org/10.1016/j.tics.2008.10.003

Böckle, M.-P., Brenden, A. P., Klingegård, M., Habibovic, A., & Bout, M. (2017). Sav2p: Exploring the impact of an interface for shared automated vehicles on pedestrians' experience. In S. Boll, A. Löcken, R. Schroeter, M. Baumann, I. Alvarez, L. Chuang, S. Feuerstack, M. Jeon, N. Broy, H. H. van Huysduynen, S. Osswald, I. Politis, & D. Large (Eds.), *Proceedings of the 9th international conference on automotive user interfaces and interactive vehicular applications adjunct* (pp. 136–140). ACM. https://doi.org/10.1145/3131726.3131765

Borrego, A., Latorre, J., Llorens, R., Alcañiz, M., & Noé, E. (2016). Feasibility of a walking virtual reality system for rehabilitation: Objective and subjective parameters. *Journal of NeuroEngineering and Rehabilitation*, *13*(1). https://doi.org/10.1186/s12984-016-0174-1

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*(4), 1061–1071. https://doi.org/10.1037/0033-295x.111.4.1061

Botvinick, M., & Cohen, J. (1998). Rubber hands 'feel' touch that eyes see. *Nature*, *391*(6669), 756–756. https://doi.org/10.1038/35784

Bown, J., White, E., & Boopalan, A. (2017). Looking for the ultimate display: A brief history of virtual reality. In J. Gackenbach & J. Bown (Eds.), *Boundaries of self and reality online* (pp. 239–259). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-804157-4.00012-8

Brewster, D. (1856). *The stereoscope; its history, theory, and construction: With its application to the fine and useful arts and to education*. John Murray.

Brosseau, M., Zangenehpour, S., Saunier, N., & Miranda-Moreno, L. (2013). The impact of waiting time and other factors on dangerous pedestrian crossings and violations at signalized intersections: A case study in Montreal. *Transportation Research Part F: Traffic Psychology and Behaviour*, *21*, 159–172. https://doi.org/https://doi.org/10.1016/j.trf.2013.09.010

Burns, C. G., Oliveira, L., Hung, V., Thomas, P., & Birrell, S. (2020). Pedestrian attitudes to shared-space interactions with autonomous vehicles — A virtual reality study. In N. Stanton (Ed.), *Advances in human factors of transportation* (pp. 307–316). Springer International Publishing.

Camara, F., Dickinson, P., & Fox, C. (2021). Evaluating pedestrian interaction preferences with a game theoretic autonomous vehicle in virtual reality. *Transportation Research Part F: Traffic Psychology and Behaviour*, *78*, 410–423. https://doi.org/10.1016/j.trf.2021.02.017

Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, *54*(4), 297–312. https://doi.org/10.1037/h0040950

Cavallo, V., Dommes, A., Dang, N.-T., & Vienne, F. (2019). A street-crossing simulator for studying and training pedestrians. *Transportation Research Part F: Traffic Psychology and Behaviour*, *61*, 217–228. https://doi.org/https://doi.org/10.1016/j.trf.2017.04.012

Chang, C.-M., Toda, K., Sakamoto, D., & Igarashi, T. (2017). Eyes on a car: An interface design for communication between an autonomous car and a pedestrian. In S. Boll, B. Pfleging, B. Donmez, I. Politis, & D. Large (Eds.), *Proceedings of the 9th international conference on automotive user interfaces and interactive vehicular applications* (pp. 65–73). ACM. https://doi.org/10.1145/3122986.3122989

Cloutier, M.-S., Lachapelle, U., d'Amours-Ouellet, A.-A., Bergeron, J., Lord, S., & Torres, J. (2017). "Outta my way!" individual and environmental correlates of interactions between pedestrians and vehicles during street crossings. *Accident Analysis & Prevention*, *104*, 36–45. https://doi.org/https://doi.org/10.1016/j.aap.2017.04.015

Colley, M., Walch, M., & Rukzio, E. (2020). Unveiling the lack of scalability in research on external communication of autonomous vehicles. *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–9. https://doi.org/10.1145/3334480.3382865

Comeau, C. P., & Bryan, J. S. (1961). Headsight television system provides remote surveillance. *Electronics*, 86–90.

Connelly, M. L., Conaglen, H. M., Parsonson, B. S., & Isler, R. B. (1998). Child pedestrians' crossing gap thresholds. *Accident Analysis & Prevention*, *30*(4), 443–453. https://doi.org/https://doi.org/10.1016/S0001-4575(97)00109-7

Corbett, M. R., & Morrongiello, B. A. (2017). Examining how different measurement approaches impact safety outcomes in child pedestrian research: Implications for research and prevention. *Accident; analysis and prevention*, *106*, 297–304. https://doi.org/10.1016/j.aap.2017.06.002

Creagh, H. (2003). Cave automatic virtual environment. *Proceedings: Electrical Insulation Conference and Electrical Manufacturing and Coil Winding Technology Conference (Cat. No.03CH37480)*, 499–504. https://doi.org/10.1109/EICEMC.2003.1247937

Cruz-Neira, C., Sandin, D., & DeFanti, T. (1993). Surround-screen projection-based virtual reality: The design and implementation of the CAVE. *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, 135–142.

Cruz-Neira, C., Sandin, D., DeFanti, T., Kenyon, R., & Hart, J. (1992). The CAVE: Audio visual experience automatic virtual environment. *Commun. ACM*, *35*(6), 64–72. https://doi.org/10.1145/129888.129892

Davis, S., Nesbitt, K., & Nalivaiko, E. (2014). A systematic review of cybersickness. *Proceedings of the 2014 Conference on Interactive Entertainment*, 1–9. https://doi.org/10.1145/2677758.2677780

Deb, S., Carruth, D. W., & Hudson, C. R. (2020). How communicating features can help pedestrian safety in the presence of self-driving vehicles: Virtual reality experiment. *IEEE Transactions on Human-Machine Systems*, *50*(2), 176–186. https://doi.org/10.1109/THMS.2019.2960517

Deb, S., Carruth, D. W., Sween, R., Strawderman, L., & Garrison, T. M. (2017). Efficacy of virtual reality in pedestrian safety research. *Applied Ergonomics*, *65*, 449–460. https://doi.org/https://doi.org/10.1016/j.apergo.2017.03.007

Deb, S., Strawderman, L. J., & Carruth, D. W. (2018). Investigating pedestrian suggestions for external features on fully autonomous vehicles: A virtual reality experiment. *Transportation Research Part F: Traffic Psychology and Behaviour*, *59*, 135–149. https://doi.org/https://doi.org/10.1016/j.trf.2018.08.016

de Clercq, K., Dietrich, A., Velasco, J. P. N., de Winter, J., & Happee, R. (2019). External human-machine interfaces on automated vehicles: Effects on pedestrian crossing decisions. *Human Factors*, *61*(8), 1353–1370. https://doi.org/10.1177/0018720819836343

de Vignemont, F. (2011). Embodiment, ownership and disownership. *Consciousness and Cognition*, *20*(1), 82–93. https://doi.org/https://doi.org/10.1016/j.concog.2010.09.004

Dietrich, A., Willrodt, J.-H., Wagner, K., & Bengler, K. (2018). Projection-based external human machine interfaces — Enabling interaction between automated vehicles and pedestrians. In A. Kemeny, F. Colombet, F. Merienne, & S. Espié (Eds.), *Proceedings of the driving simulation conference 2018 europe vr* (pp. 43–50).

Doric, I., Frison, A.-K., Wintersberger, P., Riener, A., Wittmann, S., Zimmermann, M., & Brandmeier, T. (2016). A novel approach for researching crossing behavior and risk acceptance: The pedestrian simulator. *Adjunct Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 39–44. https://doi.org/10.1145/3004323.3004324

Dörner, R., Broll, W., Grimm, P., & Jung, B. (Eds.). (2019). *Virtual und augmented reality (VR/AR)*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-58861-1

Dużmańska, N., Strojny, P., & Strojny, A. (2018). Can simulator sickness be avoided? A review on temporal aspects of simulator sickness. *Frontiers in Psychology*, *9*, 2132. https://doi.org/10.3389/fpsyg.2018.02132

Epke, M. R., Kooijman, L., & de Winter, J. C. F. (2021). I see your gesture: A VR-based study of bidirectional communication between pedestrians and automated vehicles. *Journal of Advanced Transportation*, *2021*, 1–10. https://doi.org/10.1155/2021/5573560

Faas, S. M., Mathis, L.-A., & Baumann, M. (2020). External HMI for self-driving vehicles: Which information shall be displayed? *Transportation Research Part F: Traffic Psychology and Behaviour*, *68*, 171–186. https://doi.org/https://doi.org/10.1016/j.trf.2019.12.009

Faas, S. M., Mattes, S., Kao, A. C., & Baumann, M. (2020). Efficient paradigm to measure street-crossing onset time of pedestrians in video-based interactions with vehicles. *Information*, *11*(7). https://doi.org/10.3390/info11070360

Feldstein, I., Dietrich, A., Milinkovic, S., & Bengler, K. (2016). A pedestrian simulator for urban crossing scenarios [13th IFAC Symposium on Analysis, Design, and Evaluation ofHuman-Machine Systems HMS 2016]. *IFAC-PapersOnLine*, *49*(19), 239–244. https://doi.org/https://doi.org/10.1016/j.ifacol.2016.10.531

Feldstein, I. T., & Dyszak, G. N. (2020). Road crossing decisions in real and virtual environments: A comparative study on simulator validity. *Accident; analysis and prevention*, *137*, 105356. https://doi.org/10.1016/j.aap.2019.105356

Feng, Y. (2021). Do different types of VR influence pedestrian route choice behaviour? A comparison study of desktop VR and HMD VR. In *Extended abstracts of the 2021 CHI conference on human factors in computing systems*. Association for Computing Machinery. https://doi.org/10.1145/3411763.3451516

Fuchs, P., Moreau, G., & Guitton, P. (2019). *Virtual reality: Concepts and technologies*. CRC Press.

Fuest, T., Maier, A. S., Bellem, H., & Bengler, K. (2020). How should an automated vehicle communicate its intention to a pedestrian? — A virtual reality study. In T. Ahram, W. Karwowski, S. Pickl, & R. Taiar (Eds.), *Human systems engineering and design ii* (pp. 195–201). Springer International Publishing. https://doi.org/10.1007/978-3-030-27928-8{\textunderscore}30

Fuest, T. E. (2021). *Comparison of methods to evaluate automated vehicle's driving behavior for communicating intentions to pedestrians by using wizard of oz, virtual reality, and video setups* (Dissertation). Technische Universität München. München.

Geruschat, D. R., Deremeik, J. T., & Whited, S. S. (1999). Head-mounted displays: Are they practical for school-age children? *Journal of Visual Impairment & Blindness*, *93*(8), 485–497. https://doi.org/10.1177/0145482X9909300802

Goradia, I., Doshi, J., & Kurup, L. (2014). A review paper on oculus rift & project morpheus. *International Journal of Current Engineering and Technology*, *4*(5), 3196–3200.

Harrell, W. A., & Bereska, T. (1992). Gap acceptance by pedestrians. *Perceptual and Motor Skills*, *75*(2), 432–434. https://doi.org/10.2466/pms.1992.75.2.432

Hartmann, M., Viehweger, M., Stolz, M., Watzenig, D., Spitzer, M., & Desmet, W. (2018). “Pedestrian in the Loop”: An approach using augmented reality. *SAE Technical Paper Series*. https://doi.org/10.4271/2018-01-1053

Heilig, M. L. (1957). *Stereoscopic-television apparatus for individual use* (U.S. Patent 2955156). https://patents.google.com/patent/US2955156A

Heilig, M. L. (1961). *Sensorama simulator* (U.S. Patent 3050870). https://patents.google.com/patent/US3050870A

Holzwarth, V., Gisler, J., Hirt, C., & Kunz, A. (2021). Comparing the accuracy and precision of SteamVR tracking 2.0 and Oculus Quest 2 in a room scale setup. *International Conference on Virtual and Augmented Reality Simulations*.

Horberry, T., Osborne, R., & Young, K. (2019). Pedestrian smartphone distraction: Prevalence and potential severity. *Transportation Research Part F: Traffic Psychology and Behaviour*, *60*, 515–523. https://doi.org/https://doi.org/10.1016/j.trf.2018.11.011

Hoxie, R. E., & Rubenstein, L. Z. (1994). Are older pedestrians allowed enough time to cross intersections safely? *Journal of the American Geriatrics Society*, *42*(3), 241–244. https://doi.org/10.1111/j.1532-5415.1994.tb01745.x

HTC Corporation. (2020). *Vive Pro user guide*. Retrieved August 25, 2021, from https://dl4.htc.com/Web_materials/Manual/Vive_Pro(Enterprise)/UserGuide/VIVE_Pro_User_Guide.pdf

Hu, H., Gooch, A., Thompson, W., Smits, B., Rieser, J., & Shirley, P. (2000). Visual cues for imminent object contact in realistic virtual environments. *Proceedings Visualization 2000. VIS 2000 (Cat. No.00CH37145)*, 179–185. https://doi.org/10.1109/VISUAL.2000.885692

Hudson, C. R., Deb, S., Carruth, D. W., McGinley, J., & Frey, D. (2019). Pedestrian perception of autonomous vehicles with external interacting features. In I. L. Nunes (Ed.), *Advances in human factors and systems interaction* (pp. 33–39). Springer International Publishing. https://doi.org/10.1007/978-3-319-94334-3{\textunderscore}5

International Ergonomics Association. (2021). What is ergonomics? Retrieved June 6, 2021, from https://iea.cc/what-is-ergonomics/

Ishii, A., Suzuki, I., Sakamoto, S., Kanai, K., Takazawa, K., Doi, H., & Ochiai, Y. (2016). Graphical manipulation of human's walking direction with visual illusion. *ACM SIGGRAPH 2016 Posters*, 1–2. https://doi.org/10.1145/2945078.2945087

Kaleefathullah, A. A., Merat, N., Lee, Y. M., Eisma, Y. B., Madigan, R., Garcia, J., & de Winter, J. (2020). External human–machine interfaces can be misleading: An examination of trust development and misuse in a cave-based pedestrian simulation environment. *Human Factors*, *0*(0), 0018720820970751. https://doi.org/10.1177/0018720820970751

Kelley, T. L. (1927). Interpretation of educational measurements.

Kelly, J., Cherep, L., & Siegel, Z. (2017). Perceived space in the HTC Vive. *ACM Transactions on Applied Perception (TAP)*, *15*(1), 1–16.

Kelly, J. W., Cherep, L. A., Klesel, B., Siegel, Z. D., & George, S. (2018). Comparison of two methods for improving distance perception in virtual reality. *ACM Trans. Appl. Percept.*, *15*(2). https://doi.org/10.1145/3165285

Kelly, J. W., Cherep, L. A., Lim, A. F., Doty, T., & Gilbert, S. B. (2021). Who are virtual reality headset owners? A survey and comparison of headset owners and non-owners. *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, 687–694. https://doi.org/10.1109/VR50410.2021.00095

Kilteni, K., Groten, R., & Slater, M. (2012). The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments*, *21*(4), 373–387. https://doi.org/10.1162/PRES_a_00124

Kobayashi, W., & Yoshioka, Y. (2020). Examination on corner shape for reducing mental stress by pedestrian appearing from blind spot of intersection. In N. Stanton (Ed.), *Advances in human factors of transportation* (pp. 301–306). Springer International Publishing. https://doi.org/10.1007/978-3-030-20503-4_28

Kooijman, L., Happee, R., & de Winter, J. (2019). How do eHMIs affect pedestrians' crossing behavior? A study using a head-mounted display combined with a motion suit. *Information*, *10*(12), 386. https://doi.org/10.3390/info10120386

Kumparak, G. (2014). A brief history of Oculus. Retrieved June 3, 2021, from https://techcrunch.com/2014/03/26/a-brief-history-of-oculus/

Lee, Y. M., Uttley, J., Solernou, A., Giles, O., Romano, R., Markkula, G., & Merat, N. (2019). Investigating pedestrians' crossing behaviour during car deceleration using wireless head mounted display: An application towards the evaluation of eHMI of automated vehicles. *Proceedings of the 10th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design: driving assessment 2019*, 252–258. https://doi.org/10.17077/drivingassessment.1703

Lin, H., Werner, K. M., & Inzlicht, M. (2021). Promises and perils of experimentation: The mutual-internal-validity problem. *Perspectives on Psychological Science*, *16*(4), 854–863. https://doi.org/10.1177/1745691620974773

Lubetzky, A. V., Kelly, J., Wang, Z., Gospodarek, M., Fu, G., Sutera, J., & Hujsak, B. D. (2020). Contextual sensory integration training via head mounted display for individuals with vestibular disorders: A feasibility study. *Disability and rehabilitation. Assistive technology*, 1–11. https://doi.org/10.1080/17483107.2020.1765419

Lundgren, V. M., Habibovic, A., Andersson, J., Lagström, T., Nilsson, M., Sirkka, A., Fagerlönn, J., Fredriksson, R., Edgren, C., Krupenia, S., & Saluäär, D. (2017). Will there be new communication needs when introducing automated vehicles to the urban context? In N. A. Stanton, S. Landry, G. Di Bucchianico, & A. Vallicelli (Eds.), *Advances in human aspects of transportation* (pp. 485–497). Springer International Publishing.

Luo, H., Yang, T., Kwon, S., Zuo, M., Li, W., & Choi, I. (2020). Using virtual reality to identify and modify risky pedestrian behaviors amongst chinese children. *Traffic injury prevention*, *21*(1), 108–113. https://doi.org/10.1080/15389588.2019.1694667

Mallaro, S., Rahimian, P., O'Neal, E. E., Plumert, J. M., & Kearney, J. K. (2017). A comparison of head-mounted displays vs. large-screen displays for an interactive pedestrian simulator. *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*. https://doi.org/10.1145/3139131.3139171

Mamidipalli, S., Sisiopiku, V., Schroeder, B., & Elefteriadou, L. (2015). A review of analysis techniques and data collection methods for modeling pedestrian crossing behaviors. *Journal of Multidisciplinary Engineering Science and Technology*, *2*(2).

Manzi, J. (2012). *Uncontrolled: The surprising payoff of trial-and-error for business, politics, and society*. Basic Books (AZ).

Maruhn, P., Dietrich, A., Prasch, L., & Schneider, S. (2020). Analyzing pedestrian behavior in augmented reality — Proof of concept. *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 313–321. https://doi.org/10.1109/VR46266.2020.00051

Mazuryk, T., & Gervautz, M. (1996). Virtual reality-history, applications, technology and future.

McGreevy, M. W. (1993). Chapter 8 - Virtual reality and planetary exploration. In A. Wexelblat (Ed.), *Virtual reality* (pp. 163–197). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-745045-2.50018-0

McLellan, H. (1996). Virtual realities. *Handbook of research for educational communications and technology*, 457–487.

Meehan, M., Insko, B., Whitton, M., & Brooks, F. P. (2002). Physiological measures of presence in stressful virtual environments. *ACM Trans. Graph.*, *21*(3), 645–652. https://doi.org/10.1145/566654.566630

Meir, A., Parmet, Y., & Oron-Gilad, T. (2013). Towards understanding child-pedestrians' hazard perception abilities in a mixed reality dynamic environment. *Transportation Research Part F: Traffic Psychology and Behaviour*, *20*, 90–107. https://doi.org/https://doi.org/10.1016/j.trf.2013.05.004

Mestre, D. R. (2017). CAVE versus head-mounted displays: Ongoing thoughts. *Electronic Imaging*, *29*(3), 31–35. https://doi.org/10.2352/issn.2470-1173.2017.3.ervr-094

Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, *77*(12), 1321–1329.

Milgram, P., Takemura, H., Utsumi, A., & Kishino, F. (1995). Augmented reality: a class of displays on the reality-virtuality continuum. In H. Das (Ed.), *Telemanipulator and telepresence technologies* (pp. 282–292). SPIE. https://doi.org/10.1117/12.197321

Millard-Ball, A. (2018). Pedestrians, autonomous vehicles, and cities. *Journal of Planning Education and Research*, *38*(1), 6–12. https://doi.org/10.1177/0739456X16675674

Morrongiello, B. A., & Corbett, M. (2015). Using a virtual environment to study child pedestrian behaviours: A comparison of parents' expectations and children's street crossing behaviour. *Injury prevention : journal of the International Society for Child and Adolescent Injury Prevention*, *21*(5), 291–295. https://doi.org/10.1136/injuryprev-2014-041508

Morrongiello, B. A., Corbett, M., Milanovic, M., Pyne, S., & Vierich, R. (2014). Innovations in using virtual reality to study how children cross streets in traffic: Evidence for evasive action skills. *Injury prevention : journal of the International Society for Child and Adolescent Injury Prevention*, *21*(4), 266–270. https://doi.org/10.1136/injuryprev-2014-041357

Muhanna, M. A. (2015). Virtual reality and the CAVE: Taxonomy, interaction challenges and research directions. *Journal of King Saud University - Computer and Information Sciences*, *27*(3), 344–361. https://doi.org/https://doi.org/10.1016/j.jksuci.2014.03.023

Naser, M. M., Zulkiple, A., Al bargi, W. A., Khalifa, N. A., & Daniel, B. D. (2017). Modeling pedestrian gap crossing index under mixed traffic condition. *Journal of Safety Research*, *63*, 91–98. https://doi.org/https://doi.org/10.1016/j.jsr.2017.08.005

Niehorster, D. C., Li, L., & Lappe, M. (2017). The accuracy and precision of position and orientation tracking in the htc vive virtual reality system for scientific research. *i-Perception*, *8*(3), 2041669517708205. https://doi.org/10.1177/2041669517708205

Nilsson, N. C., Peck, T., Bruder, G., Hodgson, E., Serafin, S., Whitton, M., Steinicke, F., & Rosenberg, E. S. (2018). 15 years of research on redirected walking in immersive virtual environments. *IEEE Computer Graphics and Applications*, *38*(2), 44–56. https://doi.org/10.1109/MCG.2018.111125628

Nilsson, N. C., Serafin, S., Steinicke, F., & Nordahl, R. (2018). Natural walking in virtual reality: A review. *Comput. Entertain.*, *16*(2). https://doi.org/10.1145/3180658

Nuñez Velasco, J. P., Mun Lee, Y., Uttley, J., Solernou, A., Farah, H., van Arem, B., Hagenzieker, M., & Merat, N. (2021). Will pedestrians cross the road before an automated vehicle? the effect of drivers' attentiveness and presence on pedestrians' road crossing behavior. *Transportation Research Interdisciplinary Perspectives*, *12*, 100466. https://doi.org/10.1016/j.trip.2021.100466

OculusVR. (2017). Oculus roomscale-tips for setting up a killer VR room. Retrieved March 14, 2022, from https://www.oculus.com/blog/oculus-roomscale-tips-for-setting-up-a-killer-vr-room/

OculusVR. (2019). Introducing Oculus Quest, our first 6DoF all-in-one VR system, launching spring 2019. Retrieved March 14, 2022, from https://www.oculus.com/blog/introducing-oculus-quest-our-first-6dof-all-in-one-vr-system-launching-spring-2019/

Orlosky, J., Weber, M., Gu, Y., Sonntag, D., & Sosnovsky, S. (2015). An interactive pedestrian environment simulator for cognitive monitoring and evaluation. *Proceedings of the 20th International Conference on Intelligent User Interfaces Companion*, 57–60. https://doi.org/10.1145/2732158.2732175

Otherson, I., Conti-Kufner, A. S., Dietrich, A., Maruhn, P., & Bengler, K. (2018). Designing for automated vehicle and pedestrian communication: Perspectives on eHMIs from older and younger persons. *Proceedings of the Human Factors and Ergonomics Society Europe*, 135–148.

Pala, P., Cavallo, V., Dang, N. T., Granié, M.-A., Schneider, S., Maruhn, P., & Bengler, K. (2021a). Analysis of street-crossing behavior: Comparing a CAVE simulator and a head-mounted display among younger and older adults. *Accident Analysis & Prevention*, *152*, 106004. https://doi.org/https://doi.org/10.1016/j.aap.2021.106004

Pala, P., Cavallo, V., Dang, N. T., Granié, M.-A., Schneider, S., Maruhn, P., & Bengler, K. (2021b). Is the street-crossing behavior with a head-mounted display different from that behavior in a CAVE? A study among young adults and children. *Transportation Research Part F: Traffic Psychology and Behaviour*, *82*, 15–31. https://doi.org/https://doi.org/10.1016/j.trf.2021.07.016

Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. Basic books.

Perez, D., Hasan, M., Shen, Y., & Yang, H. (2019). AR-PED: A framework of augmented reality enabled pedestrian-in-the-loop simulation. *Simulation Modelling Practice and Theory*, *94*, 237–249. https://doi.org/https://doi.org/10.1016/j.simpat.2019.03.005

Pfeil, K., Masnadi, S., Belga, J., Sera-Josef, J.-V. T., & LaViola, J. (2021). Distance perception with a video see-through head-mounted display. In *Proceedings of the 2021 chi conference on human factors in computing systems*. Association for Computing Machinery. https://doi.org/10.1145/3411764.3445223

Plumert, J. M., Kearney, J. K., & Cremer, J. F. (2004). Children's perception of gap affordances: Bicycling across traffic-filled intersections in an immersive virtual environment. *Child development*, *75*(4), 1243–1253. https://doi.org/10.1111/j.1467-8624.2004.00736.x

Ragan, E. D., Bowman, D. A., Kopper, R., Stinson, C., Scerbo, S., & McMahan, R. P. (2015). Effects of field of view and visual complexity on virtual reality training effectiveness for a visual scanning task. *IEEE transactions on visualization and computer graphics*, *21*(7), 794–807. https://doi.org/10.1109/TVCG.2015.2403312

Rebenitsch, L., & Owen, C. (2020). Estimating cybersickness from virtual reality applications. *Virtual Reality*, *25*(1), 165–174. https://doi.org/10.1007/s10055-020-00446-6

Renner, R. S., Velichkovsky, B. M., & Helmert, J. R. (2013). The perception of egocentric distances in virtual environments - A review. *ACM Comput. Surv.*, *46*(2). https://doi.org/10.1145/2543581.2543590

Rid, T. (2016). *Rise of the machines : A cybernetic history*. W.W. Norton & Company.

Ridene, T., Leroy, L., & Chendeb, S. (2015). Innovative virtual reality application for road safety education of children in urban areas. In G. Bebis, R. Boyle, B. Parvin, D. Koracin, I. Pavlidis, R. Feris, T. McGraw, M. Elendt, R. Kopper, E. Ragan, Z. Ye, & G. Weber (Eds.), *Advances in visual computing* (pp. 797–808). Springer International Publishing.

Rodríguez Palmeiro, A., van der Kint, S., Vissers, L., Farah, H., de Winter, J., & Hagenzieker, M. (2018). Interaction between pedestrians and automated vehicles: A wizard of oz experiment. *Transportation Research Part F: Traffic Psychology and Behaviour*, *58*, 1005–1020. https://doi.org/https://doi.org/10.1016/j.trf.2018.07.020

Rojas, F., Tarnogol, F., & Yang, H. S. (2016). Dynamic social formations of pedestrian groups navigating and using public transportation in a virtual city. *The Visual Computer*, *32*(3), 335–345. https://doi.org/10.1007/s00371-015-1187-7

Rojas, F. A., & Yang, H. S. (2013). Immersive human-in-the-loop HMD evaluation of dynamic group behavior in a pedestrian crowd simulation that uses group agent-based steering. In C. C. L. Wang, N. Magnenat-Thalmann, & Z. Pan (Eds.), *Proceedings of the 12th acm siggraph international conference on virtual-reality continuum and its applications in industry - vrcai '13* (pp. 31–40). ACM Press. https://doi.org/10.1145/2534329.2534336

Rojas, F. A., Yang, H. S., & Tarnogol, F. M. (2014). Safe navigation of pedestrians in social groups in a virtual urban environment. *2014 International Conference on Cyberworlds*, 31–38. https://doi.org/10.1109/CW.2014.13

Savino, G.-L., Emanuel, N., Kowalzik, S., Kroll, F., Lange, M. C., Laudan, M., Leder, R., Liang, Z., Markhabayeva, D., Schmeißer, M., Schütz, N., Stellmacher, C., Xu, Z., Bub, K., Kluss, T., Maldonado, J., Kruijff, E., & Schöning, J. (2019). Comparing pedestrian navigation methods in virtual reality and real life. In W. Gao, H. M. Ling Meng, M. Turk, S. R. Fussell, B. Schuller, Y. Song, & K. Yu (Eds.), *2019 international conference on multimodal interaction* (pp. 16–25). ACM. https://doi.org/10.1145/3340555.3353741

Schneider, S., Ratter, M., & Bengler, K. (2019). Pedestrian behavior in virtual reality: Effects of gamification and distraction. *Proceedings of the Road Safety and Simulation Conference 2019.*

Schneider, S., & Bengler, K. (2020). Virtually the same? Analysing pedestrian behaviour by means of virtual reality. *Transportation Research Part F: Traffic Psychology and Behaviour*, *68*, 231–256. https://doi.org/https://doi.org/10.1016/j.trf.2019.11.005

Schneider, S., & Li, G. (2020). Virtual scenarios for pedestrian research: A matter of complexity? In J. Y. C. Chen & G. Fragomeni (Eds.), *Virtual, augmented and mixed reality. design and interaction* (pp. 171–190). Springer International Publishing.

Schneider, S., Maruhn, P., & Bengler, K. (2018). Locomotion, non-isometric mapping and distance perception in virtual reality, 22–26. https://doi.org/10.1145/3192975.3193022

Schneider, S., Maruhn, P., Dang, N.-T., Pala, P., Cavallo, V., & Bengler, K. (2021). Pedestrian crossing decisions in virtual environments: Behavioral validity in CAVEs and head-mounted displays. *Human factors*, 18720820987446. https://doi.org/10.1177/0018720820987446

Schneider, S. A. E. (2021). *Behavioral validity in virtual reality pedestrian simulators* (Dissertation). Technische Universität München. München.

Schuemie, M. J., van der Straaten, P., Krijn, M., & van der Mast, C. A. (2001). Research on presence in virtual reality: A survey. *CyberPsychology & Behavior*, *4*(2), 183–201. https://doi.org/10.1089/109493101300117884

Schwebel, D. C., Davis, A. L., & O'Neal, E. E. (2012). Child pedestrian injury: A review of behavioral risks and preventive strategies. *American journal of lifestyle medicine*, *6*(4), 292–302. https://doi.org/10.1177/0885066611404876

Schwebel, D. C., Severson, J., & He, Y. (2017). Using smartphone technology to deliver a virtual pedestrian environment: Usability and validation. *Virtual reality*, *21*(3), 145–152. https://doi.org/10.1007/s10055-016-0304-x

Simpson, G., Johnston, L., & Richardson, M. (2003). An investigation of road crossing in a virtual environment. *Accident Analysis & Prevention*, *35*(5), 787–796. https://doi.org/https://doi.org/10.1016/S0001-4575(02)00081-7

Slater, M. (2003). A note on presence terminology. *Presence connect*, *3*(3), 1–5.

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1535), 3549–3557. https://doi.org/10.1098/rstb.2009.0138

Slater, M., Spanlang, B., & Corominas, D. (2010). Simulating virtual environments within virtual environments as the basis for a psychophysics of presence. *ACM Trans. Graph.*, *29*(4). https://doi.org/10.1145/1778765.1778829

Smus, B., Plagemann, C., & Coz, D. (2014). Cardboard: VR for android. Retrieved March 14, 2022, from https://www.youtube.com/watch?v=DFog2gMnm44

Stadler, S., Cornet, H., & Frenkler, F. (2020). Collecting people's preferences in immersive virtual reality: A case study on public spaces in Singapore, Germany, and France. *DRS2020: Synergy*. https://doi.org/10.21606/drs.2020.308

Stadler, S., Cornet, H., Novaes Theoto, T., & Frenkler, F. (2019). A tool, not a toy: Using virtual reality to evaluate the communication between autonomous vehicles and pedestrians. In M. C. tom Dieck & T. Jung (Eds.), *Augmented reality and virtual reality: The power of ar and vr for business* (pp. 203–216). Springer International Publishing. https://doi.org/10.1007/978-3-030-06246-0_15

Statistisches Bundesamt. (2020a). Statistisches Bundesamt — Verkehrsunfälle 2019: Fachserie 8 Reihe 7. https://de.statista.com/statistik/studie/id/75631/dokument/daten-zu-verkehrsunfaellen-2019/

Statistisches Bundesamt. (2020b). Verkehrsunfälle in Deutschland: Getötete im Straßenverkehr 2019. Retrieved June 10, 2021, from https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Verkehrsunfaelle/_Grafik/_Statisch/verkehrsunfaelle-getoetete-beteiligungsart.png

Statistisches Bundesamt. (2021a). GENESIS-Online 12411-0008: Bevölkerung: Deutschland, Stichtag, Altersjahre, Nationalität, Geschlecht, Familienstand. Retrieved June 11, 2021, from https://www-genesis.destatis.de

Statistisches Bundesamt. (2021b). GENESIS-Online 46241-0011: Unfallbeteiligte, Hauptverursacher des Unfalls: 46241-0011: Unfallbeteiligte, Hauptverursacher des Unfalls: Deutschland, Jahre, Geschlecht, Altersgruppen, Art der Verkehrsbeteiligung, Unfallkategorie, Ortslage. Verfügbarer Zeitraum: 2008 - 2019. Retrieved June 11, 2021, from https://www-genesis.destatis.de

Steed, A., Frlston, S., Lopez, M. M., Drummond, J., Pan, Y., & Swapp, D. (2016). An 'in the wild' experiment on presence and embodiment using consumer virtual reality equipment. *IEEE Transactions on Visualization and Computer Graphics*, *22*(4), 1406–1414. https://doi.org/10.1109/TVCG.2016.2518135

Steinicke, F., Bruder, G., Jerald, J., Frenz, H., & Lappe, M. (2008). Analyses of human sensitivity to redirected walking. *Proceedings of the 2008 ACM Symposium on Virtual Reality Software and Technology*, 149–156. https://doi.org/10.1145/1450579.1450611

Sun, Q., Patney, A., & Steinicke, F. (2020). Redirected walking in VR. In M. Magnor & A. Sorkine-Hornung (Eds.), *Real VR — Immersive digital reality: How to import the real world into head-mounted immersive displays* (pp. 285–292). Springer International Publishing. https://doi.org/10.1007/978-3-030-41816-8_12

Sutherland, I. (1965). The ultimate display. In A. K. Wayne (Ed.). Macmillan; Co.

Sutherland, I. (1968). A head-mounted three dimensional display. *Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I*, 757–764. https://doi.org/10.1145/1476589.1476686

Sutherland, I., & Sproull, B. (1996). Virtual reality before it had that name [Catalog Number: 102639877, Computer History Museum]. https://www.computerhistory.org/collections/catalog/102639877

Sween, R., Deb, S., Carruth, D. W., Waddell, D., & Furuichi, M. (2017). Development of an effective pedestrian simulator for research. In N. A. Stanton, S. Landry, G. Di Bucchianico, & A. Vallicelli (Eds.), *Advances in human aspects of transportation* (pp. 183–191). Springer International Publishing.

te Velde, A. F., van der Kamp, J., Barela, J. A., & Savelsbergh, G. J. (2005). Visual timing and adaptive behavior in a road-crossing simulation study. *Accident Analysis & Prevention*, *37*(3), 399–406. https://doi.org/https://doi.org/10.1016/j.aap.2004.12.002

Tram, T., & Parker, C. (2020). Designing exocentric pedestrian navigation for AR head mounted displays. In R. Bernhaupt, F. '. Mueller, D. Verweij, J. Andres, J. McGrenere, A. Cockburn, I. Avellino, A. Goguey, P. Bjørn, S. Zhao, B. P. Samson, & R. Kocielnik (Eds.), *Extended abstracts of the 2020 chi conference on human factors in computing systems* (pp. 1–8). ACM. https://doi.org/10.1145/3334480.3382868

UN General Assembly. (2015). *Transforming our world : The 2030 agenda for sustainable development* [A/RES/70/1]. https://sdgs.un.org/publications/transforming-our-world-2030-agenda-sustainable-development-17981

Varjo. (2021). XR-3 & VR-3 product book.

Verdelet, G., Desoche, C., Volland, F., Farnè, A., Coudert, A., Hermann, R., Truy, E., Gaveau, V., Pavani, F., & Salemme, R. (2019). Assessing spatial and temporal reliability of the Vive system as a tool for naturalistic behavioural research. *2019 International Conference on 3D Immersion (IC3D)*, 1–8. https://doi.org/10.1109/IC3D48390.2019.8975994

Vilar, E., Rebelo, F., & Noriega, P. (2014). Indoor human wayfinding performance using vertical and horizontal signage in virtual reality. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *24*(6), 601–615. https://doi.org/10.1002/hfm.20503

Vizzari, G. (2020). Virtual reality to study pedestrian wayfinding: Motivations and an experiment on usability. *2020 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, 205–208. https://doi.org/10.1109/AIVR50618.2020.00042

Walker, F., Dey, D., Martens, M., Pfleging, B., Eggen, B., & Terken, J. (2019). Feeling-of-safety slider: Measuring pedestrian willingness to cross roads in field interactions with vehicles. *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–6. https://doi.org/10.1145/3290607.3312880

Webster, M., & Sell, J. (2014). *Laboratory experiments in the social sciences.* Elsevier Science. https://books.google.de/books?id=cGBzAwAAQBAJ

Weech, S., Kenny, S., & Barnett-Cowan, M. (2019). Presence and cybersickness in virtual reality are negatively related: A review. *Frontiers in Psychology*, *10*, 158. https://doi.org/10.3389/fpsyg.2019.00158

Weidner, F., Hoesch, A., Poeschl, S., & Broll, W. (2017). Comparing VR and non-VR driving simulations: An experimental user study. *2017 IEEE Virtual Reality (VR)*, 281–282. https://doi.org/10.1109/VR.2017.7892286

Wheatstone, C. (1838). Xviii. contributions to the physiology of vision.—part the first. on some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical transactions of the Royal Society of London*, (128), 371–394.

Wickens, C. D., & Carswell, C. M. (2021). Information processing. In *Handbook of human factors and ergonomics* (pp. 114–158). John Wiley & Sons, Ltd. https://doi.org/https://doi.org/10.1002/9781119636113.ch5

Wickens, C. D., & Flach, J. M. (1988). 5 - information processing. In E. L. Wiener & D. C. Nagel (Eds.), *Human factors in aviation* (pp. 111–155). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-08-057090-7.50011-4

Willemsen, P., Colton, M., Creem-Regehr, S., & Thompson, W. (2009). The effects of head-mounted display mechanical properties and field of view on distance judgments in virtual environments. *TAP*, *6*. https://doi.org/10.1145/1498700.1498702

Wilmut, K., & Purcell, C. (2021). Why are older adults more at risk as pedestrians? A systematic review. *Human Factors*, *0*(0), 0018720821989511. https://doi.org/10.1177/0018720821989511
Wirth, T. D., & Warren, W. H. (2021). Robust weighted averaging accounts for recruitment into collective motion in human crowds. *Frontiers in applied mathematics and statistics*, *7*. https://doi.org/10.3389/fams.2021.761445
World Health Organization. (2013). *Make walking safe: A brief overview of pedestrian safety around the world* [WHO Reference Number: WHO/NMH/VIP13.02].
World Health Organization. (2018). *Global status report on road safety 2018*.
Wynne, R. A., Beanland, V., & Salmon, P. M. (2019). Systematic review of driving simulator validation studies. *Safety Science*, *117*, 138–151. https://doi.org/https://doi.org/10.1016/j.ssci.2019.04.004
Xiong, J., Tan, G., Zhan, T., & Wu, S.-T. (2020). Breaking the field-of-view limit in augmented reality with a scanning waveguide display. *OSA Continuum*, *3*(10), 2730–2740. https://doi.org/10.1364/OSAC.400900
Ye, Y., Wong, S., Li, Y., & Lau, Y. (2020). Risks to pedestrians in traffic systems with unfamiliar driving rules: A virtual reality approach. *Accident Analysis & Prevention*, *142*, 105565. https://doi.org/https://doi.org/10.1016/j.aap.2020.105565
Yuan, Y., & Steed, A. (2010). Is the rubber hand illusion induced by immersive virtual reality? *2010 IEEE Virtual Reality Conference (VR)*, 95–102. https://doi.org/10.1109/VR.2010.5444807
Zhou, N.-N., & Deng, Y.-L. (2009). Virtual reality: A state-of-the-art survey. *International Journal of Automation and Computing*, *6*(4), 319–325.

Publications

PLOS ONE

RESEARCH ARTICLE

Measuring egocentric distance perception in virtual reality: Influence of methodologies, locomotion and translation gains

Philipp Maruhn☯*, **Sonja Schneider**☯, **Klaus Bengler**☯

Chair of Ergonomics, Department of Mechanical Engineering, Technical University of Munich, Munich, Bavaria, Germany

☯ These authors contributed equally to this work.
* philipp.maruhn@tum.de

Citation: Maruhn P, Schneider S, Bengler K (2019) Measuring egocentric distance perception in virtual reality: Influence of methodologies, locomotion and translation gains. PLoS ONE 14(10): e0224651. https://doi.org/10.1371/journal.pone.0224651

Editor: Hong-jin Sun, McMaster University, CANADA

Received: November 22, 2018

Accepted: October 19, 2019

Published: October 31, 2019

Data Availability Statement: All relevant data are available from the OSF project page (https://osf.io/pnw73/, DOI 10.17605/OSF.IO/PNW73).

Funding: This research was supported by the German Research Foundation (http://www.dfg.de/en/), project number 317326196, project "Cross-Platform Validation of Pedestrian Simulators". The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abstract

Virtual reality has become a popular means to study human behavior in a wide range of settings, including the role of pedestrians in traffic research. To understand distance perception in virtual environments is thereby crucial to the interpretation of results, as reactions to complex and dynamic traffic scenarios depend on perceptual processes allowing for the correct anticipation of future events. A number of approaches have been suggested to quantify perceived distances. While previous studies imply that the selected method influences the estimates' accuracy, it is unclear how the respective estimates depend on depth information provided by different perceptual modalities. In the present study, six methodological approaches were compared in a virtual city scenery. The respective influence of visual and non-visual cues was investigated by manipulating the ratio between visually perceived and physically walked distances. In a repeated measures design with 30 participants, significant differences between methods were observed, with the smallest error occurring for visually guided walking and verbal estimates. A linear relation emerged between the visual-to-physical ratio and the extent of underestimation, indicating that non-visual cues during walking affected distance estimates. This relationship was mainly evident for methods building on actual or imagined walking movements and verbal estimates.

Introduction

Continuous technological progress renders virtual reality (VR) applications increasingly popular. Fostered by the games industry, head-mounted displays (HMDs) steadily gain in performance. Unprecedented opportunities to design flexible and highly controllable virtual environments make this technology attractive for a broad range of scientific applications, with the investigation of human behavior being a key research area.

Due to the high interest in traffic safety research, pedestrian simulators, displaying virtual traffic scenarios from a pedestrian's perspective, constitute a common use case. During the past decade, a broad range of simulator setups has been presented, with many of the more recent ones relying on different types of HMDs. Areas of research include the interaction between different road users [1], street crossing behavior [2], or effects of distraction [3].

Since many studies target collision avoidance and thus require the correct localization of objects, understanding depth perception within this specific context is essential. To decide, for example, whether the time is sufficient to cross a street in front of an approaching vehicle, one has to accurately assess both the current distance to the vehicle and the lane width in relation to walking speed. Similarly, estimates of vehicle speed and acceleration at least partially build on an assessment of what distance was covered within a certain time, equally requiring appropriate distance perception. Previous results indicate that the perception of virtual distances cannot be assumed veridical. Collisions, for example, may result from an underestimation of vehicle speeds and an overestimation of inter-vehicular gaps [2].

Since perceptual processes cannot be observed directly, a methodological challenge consists in quantifying perceived distances. To this aim, various approaches have been suggested, ranging from verbal statements to imagined or actual movements [4, 5]. While existing approaches differ in terms of accuracy and space requirements, little is known as to whether they are equally affected by different types of perceptual cues. Since expanding tracking space renders naturalistic walking increasingly feasible, in particular effects of active locomotion seem relevant. Focusing on the influence of associated visual and non-visual cues, the present study aims to compare common methods used to quantify perceived distances.

Depth perception in virtual reality

Depth perception can be defined as the ability to perceive the volume of objects as well as their relative position in three-dimensional space [6]. Egocentric depth perception thereby refers to the space between an observer and a reference, whereas exocentric distances concern the space between two external objects. In virtual environments, egocentric distances consistently tend to be underestimated [4, 7, 8], whereas [2] reported an overestimation of exocentric distances. Underestimations in particular affect egocentric distances larger than 1.0 m [8]. A relatively constant degree of underestimation between 2.0 and 7.0 m indicates a categorical rather than a continuous increase in distance compression [7].

Regarding the multisensory integration of depth cues, most literature focuses on visual perception and its interplay with proprioceptive and vestibular feedback resulting from active motion [9, 10]. Auditory [11] and haptic cues [12], in contrast, are likely to influence depth perception to some extent, but not necessarily applicable to all virtual environments. In the following, we thus focus on visual, proprioceptive, and vestibular information.

Visual depth perception

Visual depth perception is based on structural, pictorial, and motion-induced cues [6] (cf. S1 Fig). Structural depth cues refer to physical adjustments and anatomic relations between the two human eyes, including stereopsis, accomodation, and vergence [6]. Pictorial depth cues arise from features of a two-dimensional scene, such as occlusion, shadows, relative size and height in the visual field, linear and aerial perspective, texture gradient, and the arrangement of edges [4, 6]. Motion-induced visual cues, such as looming, optic flow, and motion parallax [4, 6, 13], further facilitate distance perception if either the spectator or objects in the visual scenery move.

Visual cues in VR may differ from physical environments. For stereoscopic displays, a dissociation of accomodation and vergence arises from presenting different images to both eyes, whereas the curvature of the lenses accommodates to the distance of the display [4, 14]. A lack of details may further limit the availability of pictorial depth cues. However, even in a photorealistic virtual environment displayed by a head-mounted camera, distances were underestimated by 23% (in comparison to only 4% in real world [7]). Similarly, visualizing a reference of known length did not result in more accurate judgments [8]. Hence, distance compression cannot primarily be attributed to a lack of visual details in simplified virtual surroundings or the cognitive misrepresentation of physical units.

Effects of locomotion

Active locomotion in terms of walking interaction seems to counteract distance compression in virtual environments [15–17]. It thereby appears more effective than other measures, such as presenting participants with a real-world reference [17]. Walking experience in a virtual environment was also shown to affect subsequent distance estimates in the physical world [16]: Prior to the walking interaction, estimates in real world were almost veridical, whereas post-interaction measurements increased by approximately 10%. In [18], however, accuracy only increased for distances that were equal to or smaller than those the participants had previously walked and the calibration of depth perception seemed most effective for larger distances.

In case of locomotion, not only visual, but also proprioceptive and vestibular feedback provides information on the distance covered. Investigating effects of optic flow in the absence of non-visual motion cues, [19] noted a persistent underestimation of the simulated distances, with larger deviations occurring at a shorter duration of the simulated movement. Although humans were thus able to interpret optic flow in terms of distance traveled, estimates were biased. Comparing depth perception in virtual and physical environments, [20] found a less pronounced effect of locomotion in VR. While again, virtual motion was inferred only from optic flow, actual walking provided vestibular and proprioceptive feedback in real world, possibly resulting in a higher gain from locomotion [20]. In the absence of vestibular feedback, [21] reported their subjects to rely primarily on visual information when assessing the distance traveled in comparison to a reference. Interestingly, however, they found proprioceptive feedback from cycling movements to enhance estimates, even if incongruent with the distance indicated by vision.

To distinguish the relative impact of different sensory modalities, the ratio between visually perceived and physically traveled distance may be adjusted. For ratios of 0.5, 1.0, and 2.0, [10] observed physical motion to have a stronger impact on distance estimates than visual perception. The authors thus assumed the sensitivity to visual cues to decrease in the presence of physical motion and interpreted their results as an example of sensory capture, with interoceptive cues overriding visual perception in case of conflicting information. For ratios of 0.7, 1.0, and 1.4, in contrast, [22] found estimates for multisensory conditions to range between unisensory conditions, implying that all available information influenced depth perception. They did, however, note a dominance of cues arising from physical movements for active locomotion, whereas visual cues seemed to prevail in passive locomotion. Elaborating on the differences between active and passive movements, they assumed vestibular cues to be more influential than proprioception, suggesting a linear weighted function to account for the integration of vestibular, proprioceptive, and visual cues.

While the previous results suggest locomotion to influence perceived distances via both visual and non-visual cues, [16] found optic flow to be not only insufficient to counteract distance compression in a blind walking task, but also irrelevant when proprioceptive and vestibular feedback were available. Such discrepancies may be related to the modalities used for the presentation and reproduction of distances. [10], for example, observed that participants strongly underestimated the distance of a visual target when walking towards it blindfolded, whereas estimates were relatively accurate when the distance was not presented visually but by passive motion. If distances were only represented visually, in contrast, they were matched relatively closely when simulating optic flow without actual movements. Visual and non-visual cues thus seem to yield specific and possibly even incongruent information. The performance in estimation tasks thereby depends on the agreement of sensory modalities used for encoding and reproducing distances. Hence, although active locomotion has been demonstrated to counteract the distance compression common to virtual environments, its effectiveness may vary for different types of distance estimates.

Methodologies for measuring depth perception

Because perceptual processes cannot be observed directly, distance estimates require subjects to express a mental state formed previously. Empirical data suggests that the mode of expression affects experimental results. [23], for example, instructed participants to either indicate when they felt the location of a reference had been reached or to adjust the location of this reference to a distance traveled previously. Distances under consideration ranged from 2 to 64 m. For distances beyond 12 m, the authors reported an underestimation of the traveled distance when placing an external object, whereas distances were overestimated when participants judged the moment they reached a given location. This effect was confirmed by a similar study conducted in a non-virtual environment for distances between 8 and 32 m [20].

Reviewing empirical user studies on egocentric distance perception in VR, [4] stressed the importance to acknowledge differences between measuring methodologies. Summarizing applicable methods, they differentiated between verbal estimates, perceptual matching, and visually directed actions. [24] furthermore distinguished visually guided and visually imagined actions based on differences between blindfolded and imaginary actions.

Verbal estimates

Verbal estimates require participants to indicate the perceived distance in a familiar or visible reference unit [4]. The target can either be visible during the judgment or participants can be blindfolded [4]. While this method does not require any translational motion and is fast and convenient to use, cognitive processing, a misrepresentation of physical measurement units, and prior knowledge might confound the results [4, 7, 24]. Estimates seem to be relatively precise for short distances [4, 8], whereas underestimation is exacerbated by large distances.

Perceptual matching

In perceptual matching, the size or distance of objects is compared to a given visual reference. With regard to VR, this reference is either virtual or must be memorized [4]. The corresponding action consists in either adjusting the size or distance of the virtual object or indicating the result of a mental comparison to the reference [24]. In the case of perceptual bisection, the midpoint of a distance is indicated, thereby providing information on relative depth perception [24].

Visually guided actions

Visually guided movements include throwing, walking and reaching as well as triangulated pointing. Common to all these measures is that the target is not visible during the distance quantification. [4] reported visually directed actions to be the most frequent measure of distance perception, with blind walking being particularly common. Although fairly accurate for a broad range of distances, cognitive processes such as counting steps might bias the results if participants are supposed to indicate a distance they previously walked to. To prevent such effects, triangulation tasks require participants to walk to a designated position and to subsequently indicate the assumed location of the object by pointing or stepping towards the corresponding direction [4].

Visually imagined actions

Visually imagined actions, with timed imagined walking being the most common variant, no longer require participants to actually perform a movement, but to indicate the expected time needed to do so [4]. Again, estimates can be given while the target is visible, as well as after subjects are blindfolded [24]. Just as verbal estimates, visually imagined actions are independent of spatial restrictions. However, estimates have to be compared to individual walking speed, usually measured prior to the actual experiment. Further variance is introduced by differences in the ability to imagine the walking process [4] and uncertainty as to whether participants mentally include phases of acceleration and deceleration.

Comparison of methods

A number of studies tried to capture the differences between measuring methods. [7] assumed verbal estimates to require a conscious representation, which is subject to systematic distortion. Comparing them to blind walking, they furthermore pointed out that measuring methods might differ in their susceptibility to manipulations, e.g. because subjects payed attention to the ground texture rather than to a horizon when walking. Despite a trend towards higher and thus more accurate estimates for blind walking, however, effects were statistically insignificant. Furthermore, altering the height of the horizon appeared to equally affect both tasks.

[5] evaluated techniques suitable for experiments in limited space, including verbal estimates, timed imagined walking, blind throwing, and blind triangulated pointing. Comparing two recent consumer HMDs (Oculus Rift and HTC Vive) to real-world behavior, they found distances of 2.0, 3.0, and 4.0 m to be underestimated by an additional 17% on average in comparison to real world. However, distances were also underestimated in real world and distinctive patterns emerged for the four methods. While timed imagined walking, for example, produced severe underestimations of more than 40%, the HMDs matched real-world performance relatively well in this case. Blind throwing in VR, in contrast, showed a generally moderate underestimation, albeit scoring far from real-world performance. Additionally, blind throwing and verbal estimates suggested the most severe underestimations for the farthest distance of 4.0 m, whereas blind pointing was least accurate for the closest distance of 2.0 m.

[24] compared timed imagined walking, verbal estimates and triangulated blind walking in a real-world outdoor environment, a tiled display wall, and a CAVE. While in all environments, timed imagined walking and verbal estimates provided similar results for distances between 2.0 and 12 m, triangulated walking seemed accurate in real-world environments only. In [25], perceptual matching provided different estimates for two virtual environments, whereas results for blind walking and verbal estimates did not reach statistical significance. Comparing verbal estimates and blind walking in real world and in the HTC Vive, only verbal estimates were less accurate in VR. The authors suggested that different perceptual cues influence different types of measures and that participants' strategies depend on the task to perform. To avoid the latter effect, they recommended to inform the participants about the nature of the respective task after the object was concealed, thus a mental representation had already been formed.

While it is possible that the expectation of a particular task causes subjects to focus on specific cues, it seems just as reasonable that different tasks rely differently on available information. For example, [25] found blind walking and size judgments to be affected by walking interaction, whereas verbal estimates were not. Consequently, when required to express the perceived distance by means of a number, participants did not seem to profit from the additional information provided by locomotion. On the other hand, effects may have been attenuated by cognitive processing: Recognizing that distances before and after the walking interaction were equal, people might have been reluctant to change their initial response.

Research questions and hypotheses

Previous research demonstrates that, although there is reason to believe that active locomotion counteracts distance compression in VR [15, 17, 22], effects are not equally evident for all measuring approaches [25]. Furthermore, specific methods such as verbal estimates may profit from walking interaction in some cases rather than others [15, 25]. For the technological setup to be used, walking interaction and in particular associated non-visual cues had already been shown to affect verbal estimates [15]. Based on the suggestions by [4], our aim was to investigate whether the observed effects could be generalized to further measures of perceived distance and to clarify the respective role of visual and non-visual cues.

Six methods were compared with regard to the effects of visual and non-visual cues during active locomotion. In addition to verbal estimates, we included visually guided walking, imagined timed walking, blind triangulated pointing, and blind throwing. Visually guided walking differs from blind walking, because, although the target and the surrounding street environment disappeared, reduced visual cues were provided (cf. section Methodology). A sixth method named virtual throwing was based on the concept of blind throwing (for details cf. section Methodology). Estimation accuracy was quantified by means of an error variable based on the ratio of the estimated and the virtually displayed distance (cf. Results). Although relative judgments, referring for example to the equality of distances, may be just as important for traffic safety, our objective was to evaluate the potential of walking to counteract distance compression in virtual environments [7], corresponding to an absolute underestimation.

When comparing methodologies, researchers frequently restrict their selection to methods with limited space requirements [5, 24]. Hence, effects of locomotion are often neglected. Our aim was to evaluate whether previously found differences could be replicated in a virtual city scenario allowing naturalistic walking. A focus was thereby on the comparison of verbal estimates to alternative approaches. Unlike visually directed actions, verbal estimates are often thought to rely on a conscious, typically numerical representation [7] and generally tend to be less accurate at least in comparison to blind walking [7, 25]. If inaccuracy was actually caused by the need for a conscious numerical quantification, one would expect smaller deviations for all other, visually directed actions.

H1. Visually directed actions, which do not require a conscious numerical representation, result in a smaller estimation error than verbal estimates.

Second, we tested for differences between visually directed methods, as have for instance been observed by [5] and [24]. For simplicity and to avoid confusion due to the use of similar but non-identical terminology [4, 24], these methods are also referred to as non-verbal.

H2. The estimation error for different visually directed actions varies.

Third, we expected measuring approaches based on walking movements, such as visually directed and visually imagined walking, to produce particularly low estimation errors if participants had walked to the target previously. This assumption was based on the finding that distance estimates were most accurate if the mode of presentation corresponded to the approach used for distance quantification [10]. For walking interaction, strategies such as counting steps seem for example most helpful if they can directly be linked to the process of distance quantification.

H3. In a scenario in which participants previously walked to the target, measuring methods referring to actual or imagined walking result in a smaller estimation error than other visually directed or imagined methods.

To evaluate specific effects of visual and non-visual cues, we adjusted the ratio between the visually displayed and the physically walked distance. The distance participants walked until they reached the target was thereby scaled for a constant visual distance (cf. section Methodology). The visually displayed distance being equal, lower translation gains corresponded to longer walking, which in turn was expected to result in higher estimates. Based on the known effects of distance compression and previous results [15], we expected lower translation gains to enhance estimation accuracy.

H4. Lower translation gains result in reduced underestimation.

Finally, the effects of translation gains were individually analyzed for the different methodologies. Based on the same assumptions as hypothesis H3, we expected methods referring to actual or imagined walking to be influenced more strongly by non-visual cues.

H5. The estimation error for methods referring to actual or imagined walking is more strongly influenced by the use of translation gains.

Methodology

Virtual environment

For comparison purposes, the same virtual environment, built in Unity 2017.3, was used as in [15] (cf. Fig 1). The environment was modeled after a typical street in Munich, Germany, and featured a large number of pictorial depth cues (e.g. occlusion, shadows, relative size of familiar objects, textures gradients, and linear perspective represented in VR by houses, parked cars, lane markings, etc.). Seven unique walking tracks were inserted into the virtual environment, and one of them was solely used for practice trials. The six experimental tracks varied in length from 3.0 to 3.5 m (in steps of 0.1 m), and the practice track covered a distance of 4.0 m.

For each trial, the environment was visible for 15 seconds. Afterwards, the street environment disappeared with just a unicolor gray ground layer and the sky remaining. Depending on the measuring approach, additional assisting content was displayed (cf. section Measuring methods).

When a new trial started, subjects were always standing on the positional marker, with the target marker aligned within their sagittal plane. Therefore, no head rotation was needed in order to estimate the distance between the positional and the target marker on the floor. According to the experimental condition, the participants either walked to the target marker and back to the positional marker or remained on the positional marker before estimating the distance. To avoid distraction, no auditory cues were presented. Although the latter can provide distance information [11], this choice seemed justified as the visual scene contained no further traffic participants or other moving objects which would typically emit sounds.

Equipment

The virtual environment was displayed on an HTC Vive HMD, featuring a dual AMOLED screen with a resolution of 1080 x 1200 pixels per eye and a field of view of 110 degrees. The

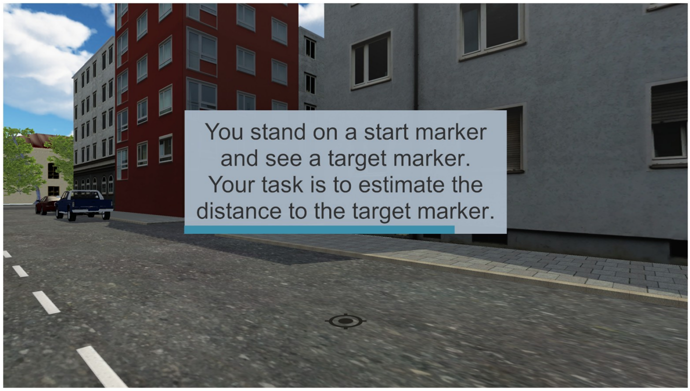

Fig 1. Virtual environment. The virtual environment replicates a typical Munich city street scenario. The environment is rich in pictorial depth cues (e.g. occlusion, shadows, relative size of familiar objects, textures gradients, and linear perspective). The scene shows a target marker to which the participants had to estimate the virtual distance. In the experiment, the written instructions were given in German. The progress bar at the bottom of the text box indicates the remaining time to read the content.

https://doi.org/10.1371/journal.pone.0224651.g001

original Chaperone system (visualization of play area boundaries) was replaced by an individual safety mesh system fitted to the experimental room. Thus, whenever the participant approached the limit of the play area (cf. S2 Fig) by less than 50 centimeters, a blue mesh was faded in to avoid collisions with physical objects in the room. The mesh was faded off as soon as the distance to all boundaries was greater than 50 centimeters again. To improve comfort and fit, the HTC Deluxe Audio Strap was used to attach the HMD to the participant's head. However, due to the lack of auditory cues, no headphones were used during the experiment. The HMD had a wired connection with a cable length of 5 m (plus 1 m from Link Box to PC). During the trials, participants held one of the HTC Vive controllers to enter and confirm distance estimates depending on the experimental condition.

The virtual environment was hosted on a VR gaming PC running on a Intel(R) Core(TM) i7 8700k CPU with 32 GB Ram and a GeForce GTX 1080 Ti graphics card. Since there were no dynamic virtual objects in the scenario, a stable frame rate with a minimum of 60 frames per second was achieved.

The room (cf. S2 Fig) allowed a maximum walking distance of 6.5 m. The experimenter was positioned in one corner of the room. By positioning the PC closer to the center of the play area, it was possible to make optimal use of the HMD's limited cable length. Two base stations (Valve lighthouse tracking) were used to track the position and rotation of the headset and controller. Contrary to the manufacturer's recommendation, the lighthouses (connected via sync cable) were set up with a distance of approx. 6.9 m. However, this did not cause any tracking problems.

Measuring methods

After being exposed to the virtual environment, the participant had to express the perceived distance to the target marker. Six different measuring methods were compared. Besides verbal estimates, these included visually directed actions as well as timed imagined walking. The

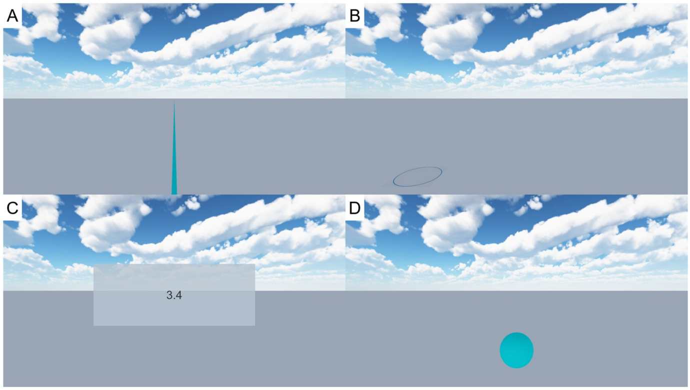

Fig 2. Measuring methods. Depending on the measuring approach, additional content was displayed after the environment was switched off. (A) An orientation line was displayed to guide the direction during visually guided walking. (B) The participant was asked to step into the circle for blind triangulated pointing. (C) A GUI displayed the participant's verbal estimate after it was entered via a keyboard by the experimenter. (D) The virtual ball the participant had to throw for virtual throwing was also visible during the flight.

https://doi.org/10.1371/journal.pone.0224651.g002

following paragraphs describe the various levels of the factor method. Specific user interfaces are depicted in Fig 2.

Verbal estimates. Verbal estimates required the subjects to explicitly verbalize the distance between the positional and the previously viewed target marker in meters with an accuracy of one tenth of a meter (e.g. 3.4 or 5.0 m). Neither the virtual environment nor the target was visible when estimates were given. The experimenter typed in the estimate, which then was displayed in the virtual environment to prevent errors due to miscommunication. If participants acknowledged the input, the experimenter confirmed it.

Visually guided walking. The visual reference was turned off and the participant was asked to walk the estimated distance [4]. Instead of blindfolding the participants, in this study, all virtual objects disappeared with only the horizon remaining, separating a unicolor gray floor from a generic sky box. In addition, an orientation line of infinite length was displayed in contrast to studies with auditory cues to ensure straight walking [26] and the participants still experienced optical flow during walking. After walking the estimated distance, subjects confirmed by pulling the controller trigger. The safety mesh was active during the whole experiment. Thus, strong overestimations triggered the safety mesh, potentially serving as a visual reference in those cases. Even for the greatest distance (3.5 m) and the smallest translation gain (0.8, cf. section Locomotion and translation gains), however, the safety mesh would only be visible at an overestimation of 40% onward.

Imagined timed walking. Participants had to imagine walking to the target (one direction). They were instructed to hold the trigger button for the duration of the imagined walk. This time measurement was then multiplied by the individual walking speed measured prior to the experiment to calculate the estimated distance.

Typically, the individual walking speed is determined by asking the participants to walk a certain distance, often several times and possibly instructing subjects to walk at a comfortable pace [5, 26–28]. Since in this study, distance perception was investigated solely in VR, the same was true for the measurement of individual walking speed. Therefore, participants were instructed to cross a visible target line at a comfortable walking speed. Walking speed was calculated as the average of four trials. The four measurements comprised two different distances, 3.0 and 3.5 m, representing the minimum and maximum distances employed in the present experiment when neglecting the translation gain. Each distance was walked once in both directions while the virtual environment was visible. The time was automatically started and stopped as soon as the participant crossed the invisible trigger boxes. The visible cyan-colored lines were 0.2 m from the trigger boxes to ensure that participants crossed both measurement points.

Blind triangulated pointing. Participants were instructed to look at the target, step to the side into a circle depicted on the floor, and point the controller towards the target. Out of the controller's lateral translation (Δx) and yaw (γ), the intercept (i) between the pointed line and the original line between participant and target can be calculated by Eq 1.

$$i = \tan(\frac{\gamma * \pi}{180^{\circ}}) * \Delta x \tag{1}$$

Various adaptations have been reported for blind triangulated pointing. [24] instructed participants to look at the target, turn by 90°, look again at the target, close their eyes, walk 2.5 m and then point to the assumed position of the target. [5] and [29] exposed participants to the target, then asked them to close their eyes, take two steps to the left and point at the target without visual feedback about the pointing.

In order to control the lateral translation in the present experiment, participants were asked to step into a circle displayed on the floor 1.5 m to the left of the original position. They then pointed at the imagined position of the target marker, confirmed by pulling the controller's trigger and returned to the original position again marked by a blue circle. There was no visual feedback indicating the pointing direction.

Blind throwing. Blind throwing was implemented according to [5]. Participants threw a bean bag (120 g) to the assumed position of the target marker. Velcro attached to the bean bag prevented it from rolling after the first contact with the carpeted floor. The experimenter measured the center position of the bean bag with the help of a second HTC Vive controller. Participants were instructed to throw from below with their strong hand. The estimation was then calculated as the longitudinal and lateral distances between the subject's position and the measured point.

Virtual throwing. Virtual throwing was developed based on the idea of blind throwing, but is independent of an experimenter returning the thrown object and allows for visual feedback of the throwing process in VR. Participants had to pick up a virtual ball with a diameter of 20 centimeters by bringing the controller to the position of the virtual object (floating in front of the participant) and pull the trigger. The ball was attached to the controller as long as the trigger was continuously pulled. With the ball still attached, the participant had to mimic a throwing movement and release the trigger button at the end of the movement to release the ball from the controller. As for blind throwing, participants were instructed to throw from below with their strong hand. The average velocity of the last five frames was passed to the ball at the point of release. SteamVR plugin for Unity—v1.2.3 (Velocity Estimator, Interactable and Throwable) was used for this purpose. On release, the ball was affected by Unity's standard gravity -9.8 world units (meter) per second squared, aerodynamic drag was ignored. The participants were instructed that the ball would not roll on the floor, thus the first point of contact between ball and virtual ground would be used as measurement. The ball was visible in mid air. In contrast to the other non-verbal methods, participants therefore received feedback regarding their performance, i.e., how far they had thrown the ball. Still, there was no feedback concerning the difference to the actual distance.

Locomotion and translation gains

In most trials, participants were instructed to once walk to the target marker and return while the environment was visible. The additional visual and non-visual depth cues were thereby subject to the systematic employment of translation gains, altering the ratio between the physical and the virtual distance. Translation gains ranged from 0.8 to 1.2 in intervals of 0.1. A translation gain of 1.0 represents an isometric mapping, where 1.0 m of physical traveled distance is experienced as 1.0 m virtual traveled. For a translation gain of 1.2, in contrast, 1.2 m were virtually passed when physically covering 1.0 m, and for translation gains smaller 1.0, participants had to walk more than 1.0 m to cover that distance in VR. Notably, the length of the experimental tracks ranging from 3.0 to 3.5 m refers to the visually displayed distance, whereas the distance to be walked was adjusted.

A translation gain greater than 1.0 can allow participants to cover longer distances in the virtual environment than possible within the restricted physical space of the experimental room and thus to explore large immersive virtual worlds. While subjective feedback in [30] indicated approval of this method, the gain was of a magnitude that rendered the difference to natural walking obvious. In the current study, in contrast, translation gains were selected to be more subtle in order to influence participants' depth perception beyond their conscious awareness.

In one trial per method, participants were instructed to experience the virtual environment from a static viewpoint instead of walking to the target marker. They thus had to rely on static visual feedback to estimate the distance, because no information from optic flow, vestibular, or proprioceptive feedback was available specifically with regard to the distance at hand. In contrast to previous studies [15, 25], however, all trials were preceded by walking movements in VR, so general scaling effects are expected to apply to all of them.

Participants

Thirty university students (age mean 26.3, SD 3.6 years) with an equal distribution of males and females were recruited. Participation required normal or corrected-to-normal vision and no prior experience in a study investigating depth perception in VR. Participants were not reimbursed in any form. They were asked about possible visual impairments and instructed to wear any visual aids under the VR glasses.

Experimental procedure

S3 Fig summarizes the study protocol as a flow chart. After providing informed consent, the subjects' interpupillary distance (IPD) was measured. For this purpose, subjects centered a measuring template on their nose from which the experimenter read the IPD. After answering demographic questions referring to gender, age, height, and visual impairments, the subject put on the HMD and adjusted the straps for a firm yet comfortable fit. The IPD of the glasses was adjusted according to the previously measured distance. All further instructions were given via text content windows, superimposed on the virtual environment.

At the beginning, subjects were instructed to walk around for two minutes and get familiar with the virtual environment. To avoid collisions with physical obstacles or walls, the area available for walking was surrounded by the virtual mesh described in the section Equipment, fading in whenever the subject approached the boundaries of the walkable area. To measure the individual walking speed, subjects were asked to cross two virtual lines, displayed on the virtual ground, at a comfortable pace. This procedure was repeated twice for two distances (3.0 and 3.5 m), resulting in a total of four measurements.

During the practice phase, subjects were exposed to the practice track six times. They were virtually translated to a positional marker and rotated towards the target marker. For half of the practice trials, subjects were instructed to walk to the target marker and back, whereas in the remaining trials, they were instructed to solely rely on static visual information in order to estimate the distance to the target marker. In all cases, the environment was visible for 15 seconds. Afterwards, one of the six measuring methods was repeated three times without displaying new virtual content in between. The repetition was chosen after participants in a pre-test had expressed the wish to practice both virtual throwing and blind throwing and it was extended to all methods for comparability and to ensure that instructions were understood. No target markers were displayed during this process in the practice phase or in the trials. The virtual scene was faded black every time the environment was toggled or the subject was re-positioned and then faded in to avoid simulator sickness. During practice, the order of measuring methods and walking interaction was constant for all 30 participants.

Subsequent to the practice phase, subjects had the opportunity to ask questions. They proceeded to the experimental phase by pushing the controller's trigger button. In contrast to the practice phase, only one distance estimate was given per trial. For each method, six trials were presented, including the five different translation gains and one trial without locomotion. Measuring methods were presented block-wise, with each of the six methods featuring six trials, resulting in 36 trials per participant. Each of the trials within a method corresponded to a different walking track, whose order was randomized. The order of methods was pseudo-randomized across participants to ensure that all possible combinations of methods for the first two positions were realized. Furthermore, the six trials within a method were pseudo-randomized in a way ensuring that for one participant, each method started with a different condition (i.e., one of the five translation gains or the non-walking condition). Completion of the experiment took 30 to 40 minutes, with participants spending approximately 25 to 35 minutes in VR.

The study design aimed to minimize the experimenter's influence, thus maximizing objectivity. All instructions in the virtual environment were given via text interfaces. Still, on some occasions, the experimenter was consulted to further clarify the instructions.

This study design was approved by the ethics committee of the Technical University of Munich (TUM School of Medicine).

Results

This study was pre-registered with the Open Science Framework (https://osf.io/69skh/). Methods used for data analysis and inferential statistics were thus predefined with minimal adjustments according to the data and review process.

Inferential statistical tests were carried out using SPSS Version 24 [31] and RStudio [32]. Table 1 gives an overview of the results regarding the hypotheses outlined in the section Research questions and hypotheses.

In all statistical analyses, estimation error (Eq 2) as a measure of accuracy served as a dependent variable. Overestimation is indicated by negative values and underestimation by positive values. As underestimation appears to be the primary problem in VR, this corresponds to expecting a numerical reduction of the estimation error when distances are perceived more

Table 1. Results overview.

	Hypothesis	Results	Results without Blind Triangulated Pointing[1]
H1	Visually directed actions, which do not require a conscious numerical representation, result in a smaller estimation error than verbal estimates.	No effect	Lower error for verbal estimates
H2	The estimation error for different visually directed actions varies.	Lower error for visually guided walking compared to blind throwing, imagined timed walking, and virtual throwing	
H3	In a scenario in which participants previously walked to the target, measuring methods referring to actual or imagined walking result in a smaller estimation error than other visually directed or imagined methods.	No effect	Lower error for walking related methods
H4	Lower translation gains result in reduced underestimation.	Reduced underestimation for lower translation gains	
H5	The estimation error for methods referring to actual or imagined walking is more strongly influenced by the use of translation gains.	Significant linear trend for all methods apart from blind triangulated pointing and virtual throwing; largest effect size for walking-related methods	

[1] Due to relatively large variances of blind triangulated pointing, analysis (in addition to pre-registered tests) has been carried out on a subset excluding these trials for all hypotheses except H5, in which methods were analyzed separately.

https://doi.org/10.1371/journal.pone.0224651.t001

accurately.

$$Estimation\ Error = 1 - \frac{Estimated\ Distance}{Visually\ Displayed\ Distance} \tag{2}$$

Walking speed

Fig 3 illustrates the walking speed recorded prior to the experiment for each of the four measurements and on average as well as the walking speed measured during the experiment. The overall mean of the four initial walking speed measurements was 0.95 m/s^2 with a standard deviation of 0.15 m/s^2.

For comparison reasons, walking speed was also measured when participants walked to the target marker during the experiment. For a given walking speed in the physical room, the virtual walking speed, i.e. the speed at which movement was displayed in the virtual environment, was thereby affected by the translation gain. Analogously to the a priori speed measurement, the walking speed measurements during the experiment featured a threshold of 0.2 m, i.e., the first and last 20 cm were not taken into account. As in some cases participants did not walk all the way to the target marker but instead stopped before the 0.2 m threshold, only 783 out of 900 walking trials were analyzed. Compared to the initial walking speed measurement, similar values were obtained for the physical (mean 0.90 m/s^2, SD 0.2 m/s^2) and the virtual speed (mean 0.92 m/s^2, SD 0.35 m/s^2) during the experiment. The larger dispersion in virtual speed indicates that participants maintained their natural walking speed when translation gains were applied, which resulted in a broader variance for virtual speed as it was affected by the translation gain.

Outliers and data exclusion

Twelve data points concerning nine test subjects were excluded due to technical errors during virtual throwing or the misunderstanding of instructions (participants triggered the controller in blind triangulated pointing before stepping into the circle or did not walk to the target when expected). These participants were excluded from analyses pertaining to the respective data subsets, leaving 21 subjects for the analysis of hypotheses H1 and H2, and 22 subjects in the

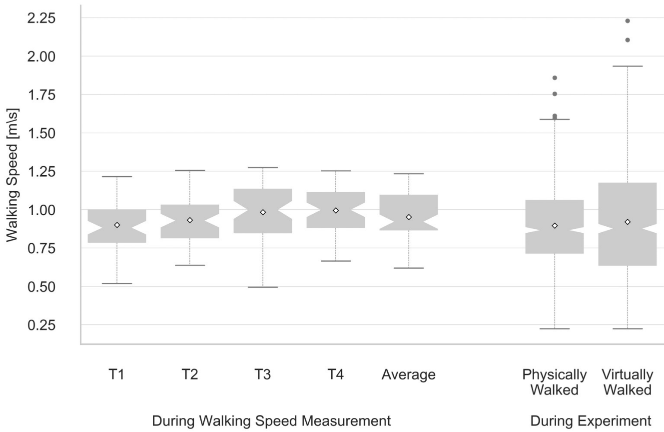

Fig 3. Walking speed results. 'T1' to 'T4' represent the results of the four walking speed measurements, 'Average' the arithmetic means of these trials. 'Physically Walked' represents the walking speed during the experiment (when participants walked to the target while the environment was visible), and 'Virtually Walked' is the virtually traveled distance. Mean values are represented by diamonds, outliers by filled circles. Notches indicate the 95% interval of the median. The whiskers show the range of the data and extend up to 1.5 of the interquartile range.

https://doi.org/10.1371/journal.pone.0224651.g003

case of H3, H4 (cf. section Research questions and hypotheses). For H5, in which methods were analyzed separately, the sample size varied between 26 and 30.

Especially for blind triangulated pointing, extreme outliers were observed (cf. S4 Fig), mainly related to one individual. Since a post experiment interview indicated no misunderstanding of instructions, the corresponding data were nonetheless included in the analysis. To simplify graphical interpretation, however, the following figures do not contain outliers. Graphs were created using Seaborn [33] for Python, treating all data points lying more than 1.5 times the interquartile range from the lower and upper quartiles as outliers.

The relatively large variance of blind triangulated pointing (cf. Fig 4) may conceal differences between other methods in statistical analyses. Therefore and in addition to pre-registered tests, inferential statistics were also carried out excluding blind triangulated pointing. As certain participants were excluded from the original data set due to missing values for this method, the corresponding subsets featured a sample size of 26 participants for hypotheses H1 to H4.

Measuring methods

Hypotheses H1 and H2 concerned differences between the methods, with H1 referring to differences in comparison to verbal estimates and H2 to differences between non-verbal methods. Non-walking trials and the five different translation gains were considered as a combined factor (as each trial either featured a translation gain or corresponded to a non-walking trial). Fig 5 shows values for this combined factor and methods in a factor plot. For the analysis of hypothesis H1, referring to the difference between verbal estimates and non-verbal methods, the data comprised all recorded trials. A Shapiro-Wilk test indicated significant deviations from normal distribution based on a $p < 0.05$ for all combinations featuring blind triangulated pointing as well as translation gain 0.9—virtual throwing and translation gain 0.9—blind

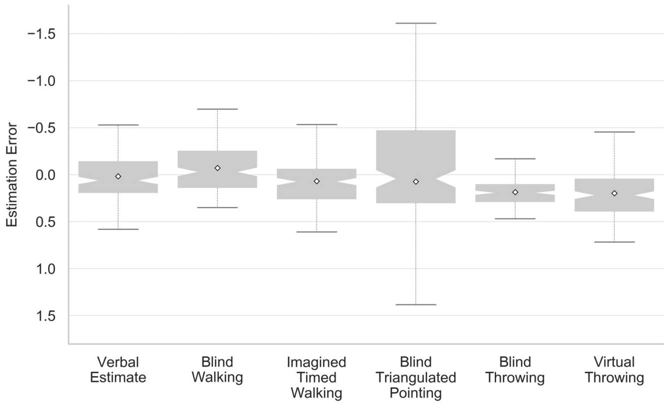

Fig 4. Estimation error for each method over all trials.

https://doi.org/10.1371/journal.pone.0224651.g004

throwing. As those groups were also part of the data set analyzed in H2 and H3, the latter were equally affected. Nonetheless, a parametric contrast analysis was performed, assigning contrast coefficients corresponding to "5" for factor combinations including verbal estimates and "-1" for all other combinations. There was no statistically significant difference in estimation error

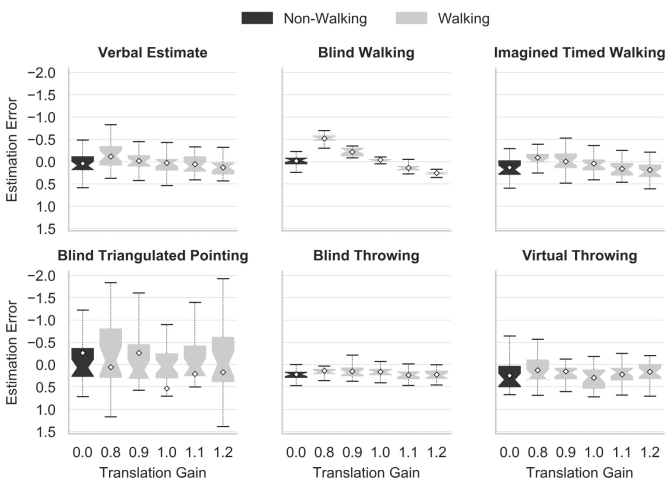

Fig 5. Translation gain factor plot over method. Factor plot displaying each method and the associated influence of translation gain. A translation gain of 0.0 corresponds to non-walking trials.

https://doi.org/10.1371/journal.pone.0224651.g005

between the verbal estimates group compared to all other groups ($F(1, 20) = 0.257$, $p = 0.618$). The contrast analysis conducted on the data subset without blind triangulated pointing trials featured coefficients of "4" for factor combinations including verbal estimates and "-1" for all other combinations. Results revealed a statistically significant difference in estimation error between the verbal estimates group compared to all other groups ($F(1, 25) = 4.735$, $p = 0.039$, $\eta^2 = 0.159$). However, in contrast to our hypothesis, verbal estimates on average seemed to result in lower estimation errors than the other methods, thus hypothesis H1 was rejected.

For hypothesis H2, all recorded trials apart from those referring to verbal estimates were included. A 5x6 two-way repeated measures ANOVA was performed including method (5 levels corresponding to five non-verbal methods) and translation gain/non-walking trials (6 levels: 0.8, 0.9, 1.0, 1.1, 1.2, non-walking). A Greenhouse-Geisser correction was applied to translation gain, method, and their interaction based on a violation of the sphericity assumption according to Mauchly's test ($p < 0.001$). The mean estimation error showed a statistically significant difference between the non-verbal methods ($F(1.137, 22.736) = 5.831$, $p = 0.021$, partial $\eta^2 = 0.226$). There was no statistically significant effect for translation gain ($F(1.665, 33.3) = 0.89$, $p = 0.403$), nor the interaction term ($F(1.689, 33.788) = 0.379$, $p = 0.653$). An ANOVA on the data subset without blind triangulated pointing, in contrast, revealed significant effects for method ($F(1.881, 47.017) = 19.934$, $p < 0.001$, partial $\eta^2 = 0.444$), translation gain ($F(3.696, 92.401) = 35.920$, $p < 0.001$, partial $\eta^2 = 0.59$) and the interaction term ($F(5.201, 130.015) = 14.095$, $p < 0.001$, partial $\eta^2 = 0.361$). Simple effects (calculated in addition to the pre-registration) thereby showed that significant differences between methods only existed for non-walking trials and translation gains smaller than 1.1 (all $p < 0.001$).

Since sphericity had been violated for method ($p < 0.001$), Bonferroni-adjusted post-hoc tests were carried out as a robust alternative to Tukey's HSD post-hoc tests. This adjustment differed from the pre-registered analysis plan. Post-hoc analysis revealed a significantly smaller estimation error ($p < 0.01$) for visually guided walking when compared to blind throwing (-.264, 95%-CI[-.336, -.192]), imagined timed walking (-.160, 95%-CI[-.283, -.036]) and virtual throwing (-.251, 95%-CI[-.341, -.161]). Post-hoc tests carried out on the data subset without blind triangulated pointing indicated the same results. No pairwise comparisons were carried out regarding translation gains, as the corresponding analysis of a linear trend was examined in the scope of hypothesis H4. Due to significant differences between non-verbal methods, hypothesis H2 was accepted.

Based on hypothesis H3, the estimation error for walking related non-verbal methods (blind and imagined timed walking) was expected to differ from the remaining non-verbal methods. Only trials in which participants walked to the target while the virtual environment was visible were considered. A Shapiro-Wilk test indicated significant deviations based on a $p < 0.05$ from normal distribution for combinations featuring a translation gain 0.9, 1.0, 1.1, 1.2—blind triangulated pointing, translation gain 0.9, 1.0—visually guided walking, translation gain 0.9—virtual throwing and translation gain 0.9—blind throwing. Still and in accordance with prior testing procedures, parametric statistical tests were chosen. A 5x5 repeated measures ANOVA with the factors translation gain (5 levels: 0.8, 0.9, 1.0, 1.1, 1.2) and method (5 levels corresponding to five non-verbal methods) in combination with a Greenhouse-Geisser correction (based on Mauchly's test values of $p < 0.001$) determined no statistically significant effect of translation gain ($F(1.202, 25.239) = 1.519$, $p = 0.234$), method ($F(1.119, 23.501) = 3.865$, $p = 0.057$), nor their interaction ($F(1.188, 24.956) = 0.340$, $p = 0.603$). Greenhouse-Geisser-corrected results based on the data set without blind triangulated pointing revealed a statistically significant effect of translation gain ($F(2.942, 73.554) = 46.950$, $p < 0.001$, partial $\eta^2 = 0.653$), method ($F(1.905, 47.623) = 19.145$, $p < 0.001$, partial $\eta^2 = 0.434$) and their interaction ($F(4.354, 108.862) = 17.38$, $p < 0.001$, partial $\eta^2 = 0.410$). Simple effects indicated that translation gains affected the estimation error for blind throwing, imagined timed walking, and visually guided walking ($p < 0.001$), but not for virtual throwing. Again, significant differences between methods were only observed for translation gains smaller than 1.1 ($p < 0.001$).

Differences between non-verbal methods related to walking were analyzed by assigning a contrast coefficient of "3" to factor combinations including visually guided walking or imagined timed walking and a coefficient of "-2" to the remaining three methods. There was no statistically significant difference in estimation error between walking-related and not walking-related non-verbal methods ($F(1, 21 = 0.140, p = 0.712$). A data subset without blind triangulated pointing was analyzed assigning a contrast coefficient of "1" to factor combinations including visually guided walking or imagined timed walking and a coefficient of "-1" to the remaining methods. Here, a statistically significant difference was found between walking-related and not walking-related methods ($F(1, 25) = 49.011$, $p < 0.001$ and $\eta^2 = 0.662$). The former generally resulted in a lower estimation error, thus supporting hypothesis H3.

Translation gains

According to hypothesis H4, lower translation gains were expected to reduce the estimation error. Just as in H3, only trials in which participants walked to the target were considered. Fig 6 illustrates the estimation error as a function of translation gain, including the non-walking condition for comparison. A linear contrast showed a statistically significant linear trend ($F(1, 21) = 8.032$, $p = 0.01$, partial $\eta^2 = 0.277$) with lower translation gains resulting in a lower estimation error, trending towards overestimation for the minimal translation gain of 0.8. The effect size increased for a linear contrast analysis of the data subset without blind triangulated pointing ($F(1, 25) = 152.938$, $p < 0.001$, partial $\eta^2 = 0.860$). Hence, hypothesis H4 was accepted.

Table 2 outlines the results concerning hypothesis H5, evaluating a possible linear trend of translation gains for each method separately. As stated in hypothesis H5, the estimation error

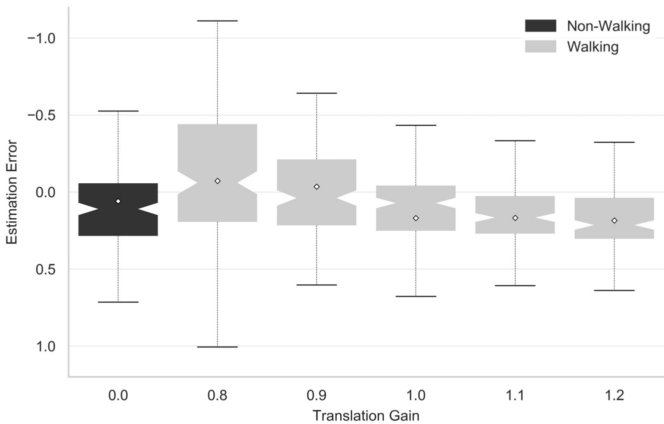

Fig 6. Overall effect of translation gain. Influence of translation gain on estimation error regardless of measuring method. A translation gain of 0.0 corresponds to non-walking trials.

https://doi.org/10.1371/journal.pone.0224651.g006

Table 2. Effect of translation gain for each method.

Method	F-value	p-value	partial η^2
Visually guided walking	F(1,28) = 1304.269	< 0.001	0.979
Imagined timed walking	F(1,27) = 141.297	< 0.001	0.840
Verbal estimate	F(1,29) = 80.457	< 0.001	0.735
Blind throwing	F(1,29) = 22.082	< 0.001	0.432
Virtual throwing	F(1,28) = 0.545	0.466	0.019
Blind triangulated pointing	F(1,25) = 0.032	0.859	0.001

https://doi.org/10.1371/journal.pone.0224651.t002

for methods referring to actual or imagined walking was most strongly influenced by the use of translation gains. Hence, hypothesis H5 was accepted. A significant linear trend, however, also produced relatively large effects for verbal estimates and blind throwing.

Exploratory analysis—Learning effects

To analyze possible learning effects, a linear mixed model allowing for intercepts to vary between participants was employed using RStudio [32] and the nlme package [34]. Predictors included the trial number within each method, the method itself (dummy coded with verbal estimates serving as a reference) and their interaction term as well as the position of the method in the experimental plan (Eq 3).

$$\begin{aligned} EstimationError \sim \quad & Trial\ Within\ Method * Method \\ & + Method\ Within\ Experiment \\ & + (1|Participant) \end{aligned} \tag{3}$$

Results (cf. S5 Fig) did not indicate statistical significance (p ≥ .155 for fixed effects) for any of the factors analyzed. Hence, no impact of learning effects across experimental trials was evident.

Discussion

The present study compared different methods to measure depth perception in VR in the context of active locomotion. The influence of visual and non-visual cues was examined by employing translation gains. Consistent with previous studies [15], an overall underestimation of distances (mean estimation error = 0.079) was observed. However, distance compression seemed less severe than in other cases [4], possibly due to technological enhancements in recent VR goggles [25]. In fact, although earlier studies indicated depth perception in real world to be almost veridical [7, 16], [25] observed an underestimation of a similar magnitude for similar distances in a physical environment.

Verbal versus non-verbal methods

As verbal estimates may be biased, for example due to an inaccurate mental representation of the reference unit, they were compared to the group of visually directed and imagined actions. In contrast to hypothesis H1, however, verbal estimates did not result in a greater underestimation (i.e., higher estimation error). In fact, the opposite seemed to be true. The mean estimation error for verbal estimates was lower than for all other methods apart from visually guided walking and overall closest to the absolute value of 0 (cf. Fig 5). A mean estimation error of approximately 2% further indicated remarkably lower underestimations in comparison to previous studies [4]. In [15], investigating distances between 3.0 and 4.0 m in the same virtual environment, the estimation error calculated according to Eq 2 averaged between 0.11 and 0.29 depending on the experimental condition. While this could be an effect of the present sample, consisting of individuals with above average capacities, similar demographic characteristics as in [15] render this explanation unlikely. Instead, presenting the distance for a relatively long time might have prevented spontaneous but inaccurate decisions. Alternatively, previous experience with other methods possibly induced carryover effects, which might be examined in future studies.

Importantly, rejecting hypothesis H1 does not contradict findings of higher consistency between real world and VR for blind walking in comparison to verbal estimates [25]. First, the experiment was conducted exclusively in VR, thus no conclusion can be drawn regarding deviations from non-virtual environments. Second, the comparison referred to the group of non-verbal methods and not to any specific approach among them. Nonetheless, the results do not support non-verbal methods in general to be more accurate than verbal estimates.

Virtual throwing

Virtual throwing was intended as a further development of blind throwing and marks the only method during which visual feedback was provided. During the experiment, numerous participants complained about their performance, mentioning they aimed for a shorter or longer throw. As for all methods, the practice phase comprised only three trials, resulting in a total of nine virtual throws. Although no significant learning effects were confirmed for any of the methods, a weak linear trend towards lower estimation errors can be seen for the virtual throwing averages over the number of trials within the method (S5 Fig). Hence, virtual throwing might benefit from increasing the number of practice trials and improving the behaviors to snap, release, and parse accelerations from the controller to the virtual ball. This might be particularly helpful, since this method has the potential to serve as an alternative for blind throwing which is applicable to experiments with limited physical space and independent of a human experimenter.

Blind triangulated pointing

In contrast to other studies [5, 29], a relatively large variance was observed for blind triangulated pointing. It is noteworthy that statistical significance concerning the comparison of measurement approaches appeared to be concealed by this noise. Due to the lack of a baseline such as open eyes pointing [29], individual pointing performance could not be compared and the comprehension of instructions could not be validated. However, post-experimental unstructured interviews did not indicate misunderstandings of the test protocol.

In contrast to previous research, the distance that participants had to step to the side was marked by a virtual circle on the ground. While the circle was meant to ensure equal displacement in all trials, it forced participants to look to the floor, thus possibly losing track of the position of the target. In future studies, participants may be guided to a constant position via either auditory signals or visual cues at eye level, enabling them to keep track of the target. Displaying the controller or a virtual light beam might additionally increase accuracy, but also reduce the meaningfulness of the method by providing an additional visual reference which is, for example, absent for blindfolded participants in real-world contexts.

Walking related methods

Visually guided walking produced significantly lower estimation errors (i.e. closer to 0) compared to imagined timed walking, blind throwing and virtual throwing. Interpreting the nonsignificant comparison to blind triangulated pointing was complicated by the large variance in this method. Overall, visually guided walking thus seemed to be a particularly precise measure, which was further indicated by the low variance, especially after walking interaction.

Considering the significantly higher estimation error, imagined timed walking does not seem to constitute an adequate replacement for visually guided walking. For imagined timed walking, distance estimates were calculated based on the average walking speed in VR, measured at the beginning of the experiment. While participants were instructed to walk at a comfortable pace, it is unclear if they imagined the same walking speed during the estimation task. The observed physical walking speed during the experiment, however, was even smaller. If participants used this value as a reference, it would thus increase the extent of underestimation. While the cognitive processes involved when imagining ego motion might differ across participants [4], manipulating the translation gain when walking in VR likely affected perceived walking speed. Nonetheless, and perhaps as a result of previous walking, estimates for imagined timed walking were notably more accurate than in [5], who employed distances between 2.0 and 4.0 m.

Effects of locomotion and translation gains

Neglecting blind triangulated pointing with regard to hypothesis H3, imagined timed walking and visually guided walking resulted in a lower estimation error than the remaining non-verbal methods (i.e., blind throwing and virtual throwing). Graphical inspection and the analysis of hypothesis H5 furthermore suggest both visually guided walking and imagined timed walking to be particularly susceptible to translation gains (cf. Fig 5), pointing to a higher impact of locomotion on walking-related tasks. The adjustment in visually guided walking, however, seems to be considerably more pronounced than for any other method.

The overall linear relationship between translation gain and estimation error was confirmed in line with [15] and according to hypothesis H4, indicating higher estimated distances for lower translation gains. While this finding underlines the effectiveness of non-visual cues arising from ego motion to influence distance estimation, the effect does not seem to affect all methods equally, as indicated by a significant interaction between method and translation gain. While no linear trend was found for blind triangulated pointing and virtual throwing, effect sizes among the remaining methods differed, with effects being particularly pronounced for visually guided walking and large but slightly less obvious for imagined timed walking and verbal estimates. For visually guided walking, the effect of translation gains below 0 even seemed to cause overestimations exceeding the adjustment in walking distance: For a translation gain of 0.8, participants walked 125% of the virtually displayed distance, but estimates referred to approximately 150% of it. Apparently, the mismatch between visual and non-visual depth cues caused participants to overshoot in visually guided walking, not just mimicking the previously experienced locomotion but overcompensating for the translation gain. For the highest translation gains of 1.1 and 1.2, in contrast, differences between methods were insignificant.

Limitations

The experimental room and dependence on a cable when using an HMD restricted the physical space available and thus the maximal distances to be analyzed. Results cannot be transferred to arbitrary distances, as depth perception has been shown to be influenced by the distance itself [8]. The wired connection to the HMD caused variable tension on the cable depending on the participant's position in the room, representing a potential additional cue. Similarly, the safety mesh that became visible when approaching the walls might have caused people to stop earlier, in particular for low translation gains and high estimates in visually guided walking. However, visual inspection of the data did not indicate such an effect (cf. S6 Fig).

Participants were not blindfolded before entering the room. Hence, it is possible that a cognitive representation of the available space biased estimates. The latter unfortunately was unavoidable, since, being students of the department, most participants were already acquainted with the room's dimensions. While a reduction of interpersonal variance might restrict confounding influences such as differences in the comprehension of instructions, the student sample also limits the generalizability to a relatively young age and people with a presumably high technical affinity.

The present study examined depth perception exclusively in VR, lacking a comparison between real and virtual environments. While this seems sufficient to the primary aim of investigating effects of active locomotion on different measuring approaches, it would be desirable to extend the results to real-world contexts.

Finally, the experimental design aimed to minimize participant-experimenter interaction. Despite extensive preliminary tests, occasional questions came up during the experiment. However, it can be seen as one step towards operator-free studies, which besides increasing objectivity might open new ways in participant recruitment for remote and online platforms as commercial VR technology becomes increasingly popular.

Conclusion

A lot of research has been carried out on depth perception in VR. However, continuous technological advancements, especially in display and tracking technologies, and an apparent dependence on compositional factors [4] require the constant reevaluation of existing knowledge. The present study demonstrated a varying impact of active locomotion on estimates provided by different experimental approaches. In particular, our results confirmed the expected effect of translation gains and indicated varying degrees of susceptibility for different methods. In comparison to previous studies, verbal estimates produced relatively accurate estimates. Hence, investigating the influence of exposure time to the visual target and carryover effects between methods could provide valuable insights regarding the application of this approach. Overall, the results demonstrate considerable differences between a number of non-verbal methods, highlighting the need for more research on differences between measures in general and between visually directed and imagined methods in particular.

Supporting information

S1 Fig. Overview of visual depth cues.
(TIF)

S2 Fig. Room setup during the experiment.
(TIF)

S3 Fig. Flow chart describing the experimental procedure.
(TIF)

S4 Fig. Outliers for blind triangulated pointing.
(TIF)

S5 Fig. Learning effects.
(TIF)

S6 Fig. Effect of distance (scaled by translation gain) on visually guided walking.
(TIF)

Author Contributions

Conceptualization: Philipp Maruhn, Sonja Schneider.

Data curation: Philipp Maruhn.

Formal analysis: Philipp Maruhn, Sonja Schneider.

Funding acquisition: Klaus Bengler.

Investigation: Philipp Maruhn, Sonja Schneider.

Methodology: Philipp Maruhn, Sonja Schneider.

Project administration: Sonja Schneider.

Software: Philipp Maruhn.

Supervision: Klaus Bengler.

Visualization: Philipp Maruhn.

Writing – original draft: Philipp Maruhn, Sonja Schneider.

Writing – review & editing: Philipp Maruhn, Sonja Schneider, Klaus Bengler.

References

1. Lehsing C, Feldstein IT. Urban interaction—getting vulnerable road users into driving simulation. In: Bengler K, Drüke J, Hoffmann S, Manstetten D, Neukum A, editors. UR:BAN Human Factors in Traffic. Wiesbaden: Springer Fachmedien Wiesbaden; 2018. p. 347–362.
2. Doric I, Frison AK, Wintersberger P, Riener A, Wittmann S, Zimmermann M, et al. A Novel Approach for Researching Crossing Behavior and Risk Acceptance. In: Green P, Pfleging B, Kun AL, Liang Y, Meschtscherjakov A, Fröhlich P, editors. Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications Adjunct—Automotive'UI 16. New York, New York, USA: ACM Press; 2016. p. 39–44.
3. Sobhani A, Farooq B. Impact of smartphone distraction on pedestrians' crossing behaviour: An application of head-mounted immersive virtual reality. Transportation Research Part F: Traffic Psychology and Behaviour. 2018; 58:228–241. https://doi.org/10.1016/j.trf.2018.06.020
4. Renner RS, Velichkovsky BM, Helmert JR. The perception of egocentric distances in virtual environments—A review. ACM Computing Surveys. 2013; 46(2):1–40. https://doi.org/10.1145/2543581.2543590
5. Peer A, Ponto K. Evaluating perceived distance measures in room-scale spaces using consumer-grade head mounted displays. In: 2017 IEEE Symposium on 3D User Interfaces (3DUI). IEEE; 2017. p. 83–86.
6. Watson MR, Enns JT. Depth Perception. In: Ramachandran VS, editor. Encyclopedia of human behavior. Boston: Academic Press; 2012. p. 690–696.
7. Messing R, Durgin FH. Distance Perception and the Visual Horizon in Head-Mounted Displays. ACM Transactions on Applied Perception. 2005; 2(3):234–250. https://doi.org/10.1145/1077399.1077403
8. Armbrüster C, Wolter M, Kuhlen T, Spijkers W, Fimm B. Depth perception in virtual reality: Distance estimations in peri- and extrapersonal space. Cyberpsychology & behavior: the impact of the Internet, multimedia and virtual reality on behavior and society. 2008; 11(1):9–15. https://doi.org/10.1089/cpb.2007.9935
9. Wexler M, van Boxtel JJA. Depth perception by the active observer. Trends in cognitive sciences. 2005; 9(9):431–438. https://doi.org/10.1016/j.tics.2005.06.018 PMID: 16099197
10. Harris LR, Jenkin M, Zikovitz DC. Visual and non-visual cues in the perception of linear self motion. Experimental brain research. 2000; 135(1):12–21. https://doi.org/10.1007/s002210000504 PMID: 11104123

11. Finnegan DJ. Compensating for Distance Compression in Virtual Audiovisual Environments [Doctoral thesis]. University of Bath. Bath; 2017.
12. Battaglia PW, Kersten D, Schrater PR. How haptic size sensations improve distance perception. PLoS computational biology. 2011; 7(6):e1002080. https://doi.org/10.1371/journal.pcbi.1002080 PMID: 21738457
13. Warren WH Jr, Kay BA, Zosh WD, Duchon AP, Sahuc S. Optic flow is used to control human walking. Nature Neuroscience. 2001; 4:213 EP –. https://doi.org/10.1038/84054 PMID: 11175884
14. Wann JP, Rushton S, Mon-Williams M. Natural problems for stereoscopic depth perception in virtual environments. Vision research. 1995; 35(19):2731–2736. https://doi.org/10.1016/0042-6989(95)00018-u PMID: 7483313
15. Schneider S, Maruhn P, Bengler K. Locomotion, Non-Isometric Mapping and Distance Perception in Virtual Reality. In: Association for Computing Machinery, editor. Proceedings of the 2018 10th International Conference on Computer and Automation Engineering—ICCAE 2018. New York, New York, USA: ACM Press; 2018. p. 22–26.
16. Waller D, Richardson AR. Correcting distance estimates by interacting with immersive virtual environments: effects of task and available sensory information. Journal of experimental psychology Applied. 2008; 14(1):61–72. https://doi.org/10.1037/1076-898X.14.1.61 PMID: 18377167
17. Kelly JW, Cherep LA, Klesel B, Siegel ZD, George S. Comparison of Two Methods for Improving Distance Perception in Virtual Reality. ACM Transactions on Applied Perception. 2018; 15(2):1–11.
18. Kelly JW, Hammel WW, Siegel ZD, Sjolund LA. Recalibration of perceived distance in virtual environments occurs rapidly and transfers asymmetrically across scale. IEEE transactions on visualization and computer graphics. 2014; 20(4):588–595. https://doi.org/10.1109/TVCG.2014.36 PMID: 24650986
19. Frenz H, Lappe M. Absolute travel distance from optic flow. Vision research. 2005; 45(13):1679–1692. https://doi.org/10.1016/j.visres.2004.12.019 PMID: 15792843
20. Lappe M, Frenz H. Visual estimation of travel distance during walking. Experimental brain research. 2009; 199(3-4):369–375. https://doi.org/10.1007/s00221-009-1890-6 PMID: 19533107
21. Sun HJ, Campos JL, Chan GSW. Multisensory integration in the estimation of relative path length. Experimental brain research. 2004; 154(2):246–254. https://doi.org/10.1007/s00221-003-1652-9 PMID: 14685814
22. Campos JL, Butler JS, Bülthoff HH. Multisensory integration in the estimation of walked distances. Experimental brain research. 2012; 218(4):551–565. https://doi.org/10.1007/s00221-012-3048-1 PMID: 22411581
23. Lappe M, Jenkin M, Harris LR. Travel distance estimation from visual motion by leaky path integration. Experimental brain research. 2007; 180(1):35–48. https://doi.org/10.1007/s00221-006-0835-6 PMID: 17221221
24. Klein E, Swan JE, Schmidt GS, Livingston MA, Staadt OG. Measurement Protocols for Medium-Field Distance Perception in Large-Screen Immersive Displays. In: Steed A, editor. IEEE Virtual Reality Conference, 2009. Piscataway, NJ: IEEE; 2009. p. 107–113.
25. Kelly JW, Cherep LA, Siegel ZD. Perceived space in the HTC Vive. ACM Transactions on Applied Perception. 2017; 15(1):1–16. https://doi.org/10.1145/3106155
26. Grechkin TY, Nguyen TD, Plumert JM, Cremer JF, Kearney JK. How does presentation method and measurement protocol affect distance estimation in real and virtual environments? ACM Transactions on Applied Perception. 2010; 7(4):1–18. https://doi.org/10.1145/1823738.1823744
27. Plumert JM, Kearney JK, Cremer JF, Recker K. Distance perception in real and virtual environments. ACM Transactions on Applied Perception. 2005; 2(3):216–233. https://doi.org/10.1145/1077399.1077402
28. Ziemer CJ, Plumert JM, Cremer JF, Kearney JK. Estimating distance in real and virtual environments: Does order make a difference? Attention, perception & psychophysics. 2009; 71(5):1095–1106.
29. Bruder G, Sanz FA, Olivier AH, Lecuyer A. Distance estimation in large immersive projection systems, revisited. In: 2015 IEEE Virtual Reality (VR). IEEE; 2015. p. 27–32.
30. Interrante V, Ries B, Lindquist J, Anderson L. Elucidating Factors that can Facilitate Veridical Spatial Perception in Immersive Virtual Environments. In: Sherman W, editor. IEEE Virtual Reality Conference, 2007. Piscataway, NJ: IEEE Service Center; 2007. p. 11–18.
31. IBM Corp. IBM SPSS Statistics for Windows, Version 24.0; Released 2016.
32. RStudio Team. RStudio: Integrated Development Environment for R; 2016. Available from: http://www.rstudio.com/.

33. Michael Waskom, Olga Botvinnik, Drew O'Kane, Paul Hobson, Saulius Lukauskas, David C Gemperline, et al. Mwaskom/Seaborn: V0.8.1 (September 2017); 2017.

34. Pinheiro J, Douglas B, DebRoy S, Sarkar D, R Core Team. nlme: Linear and Nonlinear Mixed Effects Models; 2018. Available from: https://CRAN.R-project.org/package=nlme.

Effects of Avatars on Street Crossing Tasks in Virtual Reality

Philipp Maruhn(✉) and Simon Hurst

Technical University of Munich, 85748 Garching, BY, Germany
philipp.maruhn@tum.de

Abstract. Head-mounted displays (HMDs) are a commonly applied tool to analyze pedestrian behavior in virtual environments. However, compared to reality, one's own body can only be represented in the form of a virtual replica. The present study examined the effects of displaying different virtual self-representations, or avatars, in a street crossing task on presence, virtual body ownership, gap acceptance and virtual collisions. 29 participants were instructed to cross a one-lane street with varying gap sizes between vehicles ranging from 1 to 6 s. Two different avatar concepts (with or without hand and finger tracking) were compared to a baseline without any visual body self-representation. Crossing was repeated ten times in each avatar condition, resulting in a total of 30 trials per participant. There was no difference in presence scores between the conditions. The illusion of virtual body ownership was stronger for an avatar that featured hands and finger tracking compared to an avatar in which only the position of the hands was displayed based on two hand-held controllers. In trials in which any avatar was present, participants accepted significantly smaller gaps to cross the street. An equal number of virtual collisions was observed for both avatars and the baseline without an avatar.

Keywords: Virtual reality · Pedestrian simulator · Avatar · Street crossing

1 Introduction

The rapid technological advancements of head-mounted displays (HMDs) led to an increasing number and variety of pedestrian simulator setups [1]. In contrast to Cave Automatic Virtual Environments (CAVEs) [2], the own body is usually invisible when wearing an HMD. Additional tracking of the extremities allows the display of a virtual self-representation in the form of an avatar [3]. However, this is often associated with additional technical and methodological efforts. Thus, HMD based pedestrian simulators often lack any form of body representation [1].

2 Avatars in Virtual Environments

Effects of avatars on perception and immersion in virtual environments have been studied from the beginning of research on the sense of presence [4]. Displaying an avatar is

N. L. Black et al. (Eds.): IEA 2021, LNNS 223, pp. 215–223, 2022.
https://doi.org/10.1007/978-3-030-74614-8_26

reported to increase spatial presence, affect an individual's behavior and trigger certain physiological responses (e.g. heart rate) [5].

Distance compression marks a commonly reported phenomenon related to virtual environments and HMDs. Especially in experiments related to pedestrian safety, a realistic distance perception is crucial. Participants tend to underestimate egocentric distances in comparison to the real world. In the context of natural walking, the feeling of not moving far enough in the virtual world relative to real movement is often reported. The presentation of a dynamic avatar seems to positively influence distance perception. However, the feeling of embodiment also appears to play a significant role. [6].

This raises the question of how the presence of avatars influences experiments such as common gap acceptance studies.

3 Methodology

3.1 Virtual Environment and Avatars

The virtual scenario was created with Unity 2018.3, modeled after a typical Munich city center environment (Fig. 1) and rendered on an HTC Vive Pro. This HMD features a resolution of 1440×1600 pixels per eye with a 110° field of view. The HMD and Leap Motion sensor (mounted on the HMD) were connected to the PC via a 6-m cable. The scene featured one way traffic on a 3.5 m wide single lane. Vehicles appeared around the corner at the end of the street and traveled at a constant speed of 30 km/h. The gap size between the vehicles increased from 1 to 6 s during the trials. The rate of increase was predefined for each trial and varied between .25 and 1 s, increasing either with each or every second car. The available physical space was sufficient to easily cross the street. Nonetheless, a blue guardian mesh was displayed for safety reasons as soon as subjects approached the physical boundaries of the room. The participants were instructed to cross the street as soon as they felt the gap size was sufficient. Once they reached the other side of the street, they turned around; the gaps of the current trial were filled with additional cars (making another safe crossing impossible) and the new set of cars for the next trial was spawned. This resulted in a seamless connection of trials and continuous traffic. In each experimental condition, participants crossed five times in each direction, resulting in ten trials per block.

The display of an avatar was varied between experimental blocks. There was either: (1) no avatar as a baseline, (2) an avatar based on Vive trackers or (3) an avatar with Vive trackers and Leap Motion tracking for the hands. In all conditions, participants were equipped with Vive trackers on both feet and the waist. For conditions (1)–(2), Vive controllers were held in both hands. The Leap Motion sensor was mounted on the HMD's front and enabled rendering of the participant's hand and finger gestures on the virtual avatar. The avatar's movements were based on an inverse kinematics model [7]. A unisex monochrome blue dummy (Fig. 1) was used as the avatar model to avoid possible mismatches related to gender and ethnicity and the associated uncanny valley effect [8].

3.2 Dependent Measures

The goal of pedestrian simulators is to observe behavior as close to reality as possible, while making use of the safety and controllability of a simulated environment. In this

Fig. 1. The virtual environment, featuring a one-way street with one lane. On the left, the avatar is visible displayed in the two conditions of Vive Controller and Leap Motion.

context, the feeling of immersing oneself in a virtual world, which is also described as presence [9], is of high interest [10]. In this study, the German version of the Igroup Presence Questionnaire (IPQ) [11] was used. The IPQ consists of 14 items that are measured on a seven-point Likert scale and assess presence in the dimensions of general presence, spatial presence, involvement, and realism.

In addition to presence, the particular focus of this study was on the examination of the subjective perception of two different avatar concepts. The German version of the Alpha IVBO [12] was used to measure the Illusion of Virtual Body Ownership (IVBO). The questionnaire features 13 items on a 7-point Likert scale and measures the dimensions Acceptance (of the virtual body), Control (feedback of motion), and Change (in self-perception).

Analyzing the gap size during street crossing tasks is an often used objective metric in pedestrian simulators [1]. For that, the positional data for cars, HMD, controllers, and trackers was continuously logged. Additionally, when crossing the middle of the street, the participant triggered a ray cast function to calculate the gap size, i.e., the difference in meters from the rear bumper of the gap opening car to the front bumper of the gap closing car. This distance was then divided by the cars speed (30 km/h) to obtain the gap size in seconds.

Besides gap sizes, virtual collisions are of core interest in street crossing simulators. This is especially relevant in this case, where research revolves around the question of whether the presence of a body representation influences the number of collisions. Invisible sphere colliders (10 cm diameter) were located on the participants virtual head, waist, hands, and feet. A collision was logged as soon as one of these triggers entered the collider encompassing each vehicle. Due to this implementation, it was possible to get hit multiple times but with different body parts (e.g., with the left foot and the waist) during the same crossing.

3.3 Study Protocol

A standardized greeting and consent form was followed by a demographics questionnaire and a measurement of the participants interpupillary distance (IDP). Three Vive trackers were attached to the feet and waist, the HMD was put on and adjusted to the individual's IDP. This was followed by a familiarization with the virtual environment including four practice road crossings without traffic and two to four crossings (until participants felt confident) with traffic featuring 6 s gaps. The participants were asked to explore their virtual body in experimental blocks in which an avatar was displayed. At the beginning of the trials, they were instructed to cross the street, after the first car had passed if they considered the gap to be large enough to do so safely. They were further informed that vehicles travelled with a constant velocity and would not brake. Each experimental block featured 10 crossings. Subsequently, the IVBO (in case an avatar was displayed) and presence questionnaire were presented in VR with the HMD on. Participants were asked to verbalize their answers before continuing with the next block after a short break to adapt the hardware for the next avatar concept and provide some rest outside the virtual environment. In total, each participant crossed the street 3×10 times, while order of conditions (No Avatar, Vive Controller and Leap Motion) was randomized. The overall duration of the experiment was approximately 30 to 40 min. This study design was approved by the university's ethics committee.

4 Results

A total of $N = 29$ subjects between 18 and 33 years ($M = 25.59$, $SD = 3.05$) participated in the study. The sample was predominantly composed of students (82.76%) with an overall gender distribution of 13 female and 16 male participants. 13 participants indicated they already had experience in VR due to gaming ($N = 5$), participation in other VR studies ($N = 4$), their own research ($N = 2$), for work ($N = 1$), or from exhibits ($N = 1$).

4.1 Presence and IVBO

Figure 2 displays the IVBO questionnaire's results. Shapiro Wilk's test indicated a deviation from the normal distribution for the differences between the Leap Motion and Vive Controller avatars in the IVBO subcategory Control. Subsequently, and to account for outliers, a Wilcoxon signed rank test was used for pairwise comparisons (Table 1). The Leap Motion avatar scored higher in all IVBO dimensions, with significant effects in Acceptance and Control.

Presence scores are summarized in Fig. 3. Shapiro Wilk's tests indicated multiple significant deviations from the normal distribution based on a $p \leq .05$ level for the groups No Avatar – General Presence, No Avatar – Spatial Presence, Vive Controller – General Presence, Leap Motion – General Presence and Leap Motion – Spatial Presence. Friedman's tests showed that there were no significant differences between the three avatar levels (No Avatar, Vive Controller and Leap Motion) for any of the presence subscales: General Presence ($X^2{}_F(2) = .419$, $p = .811$), Spatial Presence ($X^2{}_F(2) = .544$, $p = .762$), Involvement ($X^2{}_{\mathrm{F}}(2) = 5.029$, $p = .081$) and Realism ($X^2{}_F(2) = 2.490$, $p = .288$).

Table 1. Descriptive and statistical results for IVBO. For pairwise comparison, a Wilcoxon signed rank test was calculated.

	Leap Motion	Vive Controller	W	p	Rank-Biserial Correlation
Acceptance	$M = 4.54$ $SD = 1.314$	$M = 3.871$ $SD = 1.548$	334.0	0.003**	0.645
Control	$M = 6.052$ $SD = 0.683$	$M = 5.638$ $SD = 1.028$	211.0	0.006**	0.668
Change	$M = 2.344$ $SD = 1.022$	$M = 2.190$ $SD = 0.828$	163.0	0.455	0.181

** $p \leq .01$

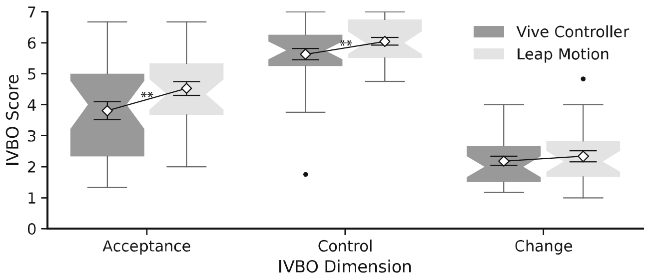

Fig. 2. IVBO scores for the Vive Controller and Leap Motion avatar. Mean values are represented by white diamonds, caps connected to the means represent the standard error. Boxplot whiskers extend up to 1.5 of the interquartile range. Notches indicate the 95% confidence interval of the median. ** $p \leq .01$

4.2 Gap Size and Virtual Collisions

Figure 4 renders gap size (in seconds) grouped by avatar. Five values were excluded since the participants crossed the street after the last experimental gap (which resulted in gap sizes > 6 s) in the respective trials. A linear mixed effects model was created in R [13] with lme4 [14]. Avatar served as fixed effect (with No Avatar as baseline) and participant as random effect (Eq.*Gap Size ~ Avatar + (1|Participant)* (1). An ANOVA with Satterthwaite's method was used to analyze the model. It revealed a significant effect of the displayed avatar on gap size when crossing the street $F(2{,}834) = 22.454$, $p < .001$ $\eta_p{}^2 = .05$. Visual inspection of QQ-plots indicated a normal distribution of residuals. Bonferroni corrected post hoc pairwise comparisons (least-squares means) revealed that smaller gaps were taken when displaying either the Vive Controller avatar ($p < .001$) or Leap Motion avatar ($p < .001$) compared to No Avatar. There was no significant difference regarding gap sizes between the two avatars ($p = .239$).

$$Gap\ Size \sim Avatar + (1|Participant) \tag{1}$$

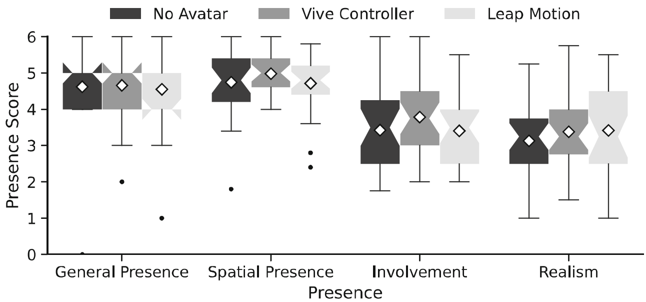

Fig. 3. IPQ Presence Scores for the three conditions No Avatar, Vive Controller and Leap Motion. The same boxplot aesthetic rules apply as in Fig. 2.

$$Gap\,Size \sim Avatar * Condition\,Order + (1|Participant) \tag{2}$$

A second model (Eq. *Gap Size ~ Avatar * Condition Order + (1|Participant)* (2) was created to account for possible learning effects. The order in which the different avatars were displayed and the interaction between order and avatar type were added as fixed effects, with the first condition as a baseline. Comparing the two models revealed that order significantly affected the gap size ($X^2(6) = 67.727$, $p < .001$), reducing the gap size by −0.53 s ($SD = 0.12$ s) for the second presented condition and by −0.29 s ($SD = 0.12$ s) for the last block. The effect size of order was ${\eta_p}^2 = .07$. Again, an ANOVA with Satterthwaite's method was used to analyze the model. There was no significant interaction ($F(4{,}847) = 1.031$, $p = .390$) between avatar and condition order. Bonferroni corrected post hoc pairwise comparisons showed a significant reduction of gap size between the first and the second trial block ($p < .001$), and the first and the third blocks ($p < .001$), but no difference between the latter two ($p = 1.0$).

Collisions were detected when any of the tracked body parts collided with a vehicle. Since the head was located too high to collide with the cars and collisions with the Leap Motion avatar's hands were not tracked, further analysis is based on the two foot and waist trackers. Overall, 245 collisions of these three trackers were recorded. 53.5% of these collisions were created by only three individuals. Lowest number of collisions per condition was recorded with the Leap Motion avatar ($M = 2.034$, $SD = 4.204$) followed by No Avatar ($M = 2.931$, $SD = 4.729$) and the Vive Controller avatar ($M = 3.483$, $SD = 6.133$) as depicted in Fig. 5. Similar distributions were observed when only the waist tracker was considered: Leap Motion avatar ($M = 1.000$, $SD = 1.871$), No Avatar ($M = 1.379$, $SD = 2.351$) and the Vive Controller avatar ($M = 1.724$, $SD = 2.776$).

Due to the deviation from normal distribution for all three levels of Avatar (based on a Shapiro Wilk test, $p < .001$), differences were analyzed with Friedman's test. There was no significant effect of the different avatar displays on the number of recorded collisions (${X^2}_F(2) = 1.680$, $p = .432$). Excluding the three individuals with the most collisions showed similar results (${X^2}_F(2) = 1.365$, $p = .505$).

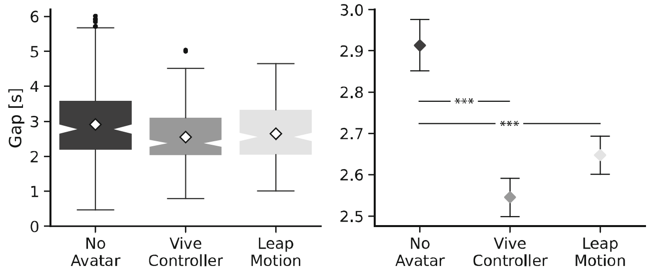

Fig. 4. Left: Overall gap acceptance values for the three conditions No Avatar, Vive Controller and Leap Motion. The same boxplot aesthetic rules apply as in Fig. 2. Right: Mean values with standard error as whiskers. For better comparability, the Y axis is limited to the relevant section. *** $p \leq .001$

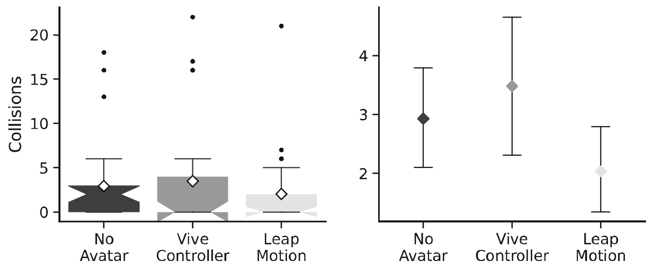

Fig. 5. Left: Number of overall collisions (sum of feet and waist tracker) in each of the three conditions No Avatar, Vive Controller and Leap Motion. The same boxplot aesthetic rules apply as in Fig. 2. Right: Mean values with standard error as whiskers. For better comparability, the Y axis is limited to the relevant section.

5 Discussion

Smaller gaps were accepted by the participants in VR when displaying avatars. Although the virtual body might increase spatial awareness and facilitate crossing decisions, spatial presence scores remained equally high for all conditions. Particularly noteworthy is the high number of collisions, which also go hand in hand with the high acceptance rates of small gaps. This also calls into question the transferability of this and other simulator studies into reality. One possible reason could be the relatively young sample of participants, that perceives VR as a game and who are therefore more likely to display risky behavior. However, it should be noted that this demographic composition is not uncommon for VR studies [1]. One might further argue that the avatar's resemblance with a crash test dummy enforced risky behavior, especially in relation to the Proteus effect [15]. This effect describes that the appearance of the avatar, for example the avatar's age or size, can influence the behavior of the user. However, due to the first-person perspective, the participants mainly experienced the avatar's hands and feet and not its overall appearance.

There are only minor differences between the two avatar conditions. The participants were encouraged to test hand gestures in the "Leap Avatar condition during familiarization, but this did not significantly affect presence scores, gap decisions or collisions during the actual experiments. Only differences in IVBO ratings compared to the avatar without hand tracking could be found. One reason might be that the hands were rarely visible when walking due to the Vive Pro's restricted field of view (110°). In other applications, where more hand interaction is required, larger differences could possibly be found.

There are also technical limitations to be mentioned: due to the leap motion controller, a wired connection to the HMD was necessary. Especially during the practice trials, participants struggled with the handling of the cable when turning around after each crossing. Whenever possible, a wireless connection should be preferred.

6 Conclusion

It could be shown that displaying an avatar impacts quantitative metrics describing street crossing behavior in terms of gap acceptance as well as subjective measures such as IVBO. However, an avatar without hand tracking seems to be sufficient for common gap acceptance tasks. Future studies should evaluate the use of additional trackers to represent an avatar.

Acknowledgments. This research was supported by the Deutsche Forschungsgemeinschaft (German Research Foundation/DFG, project number 317326196).

References

1. Schneider, S., Bengler, K.: Virtually the same? Analysing pedestrian behaviour by means of virtual reality. Transp. Res. Part F: Traff. Psychol. Behav. **68**, 231–256 (2020). https://doi.org/10.1016/j.trf.2019.11.005
2. Cavallo, V., Dommes, A., Dang, N.-T., Vienne, F.: A street-crossing simulator for studying and training pedestrians. Transp. Res. Part F: Traff. Psychol. Behav. **61**, 217–228 (2019). https://doi.org/10.1016/j.trf.2017.04.012
3. Feldstein, I., Dietrich, A., Milinkovic, S., Bengler, K.: A Pedestrian simulator for urban crossing scenarios. IFAC-PapersOnLine **49**(19), 239–244 (2016). https://doi.org/10.1016/j.ifacol.2016.10.531
4. Slater, M., Wilbur, S.: A Framework for Immersive Virtual Environments (FIVE): speculations on the role of presence in virtual environments. Presence: Teleoper. Virtual Environ. **6**(6), 603–616 (1997). https://doi.org/10.1162/pres.1997.6.6.603
5. Biocca, F.: Connected to my avatar. In: Social computing and social media: 6th international conference, SCSM 2014, held as part of HCI International 2014, Heraklion, Crete, Greece, June 22–27, 2014, proceedings, Cham, pp. 421–429 (2014)

6. Renner, R.S., Velichkovsky, B.M., Helmert, J.R.: The perception of egocentric distances in virtual environments - a review. ACM Comput. Surv. **46**(2), 1–40 (2013). https://doi.org/10.1145/2543581.2543590
7. Rootmotion, Final IK v.1.8: Rootmotion, 2018. https://root-motion.com/. Accessed 24 Jan 2021
8. Schwind, V., Wolf, K., Henze, N.: Avoiding the uncanny valley in virtual character design. Interactions **25**(5), 45–49 (2018). https://doi.org/10.1145/3236673
9. Slater, M.: Measuring presence: a response to the Witmer and singer presence questionnaire. Presence Teleoper. Virtual Environ. **8**(5), 560–565 (1999). https://doi.org/10.1162/105474699566477
10. Witmer, B.G., Singer, M.J.: Measuring presence in virtual environments: a presence questionnaire. Presence: Teleoper. Virtual Environ. **7**(3), 225–240 (1998). https://doi.org/10.1162/105474698565686.
11. Schubert, T., Friedmann, F., Regenbrecht, H.: The experience of presence: factor analytic insights. Presence: Teleoper. Virtual Environ. 10(3), 266–281 (2001). https://doi.org/10.1162/105474601300343603.
12. Roth, D., Lugrin, J.-L., Latoschik, M.E., Huber, S.: Alpha IVBO - construction of a scale to measure the illusion of virtual body ownership. In: Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, Denver Colorado USA, 2017, pp. 2875–2883 (2017)
13. R. Core Team: R: A Language and Environment for Statistical Computing. Vienna, Austria. https://www.r-project.org/
14. Bates, D., Mächler, M., Bolker, B., Walker, S.: Fitting linear mixed-effects models using lme4. J. Stat. Software **67**(1), 1–48 (2015). https://doi.org/10.18637/jss.v067.i01
15. Reinhard, R., Shah, K.G., Faust-Christmann, C.A., Lachmann, T.: Acting your avatar's age: effects of virtual reality avatar embodiment on real life walking speed. Media Psychol. **23**(2), 293–315 (2020). https://doi.org/10.1080/15213269.2019.1598435

2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)

Analyzing Pedestrian Behavior in Augmented Reality — Proof of Concept

Philipp Maruhn* André Dietrich† Lorenz Prasch‡ Sonja Schneider§

Technical University of Munich, Germany

Figure 1: A virtual car passing by, superimposed on a real street, seen through a head-mounted display.

ABSTRACT

With recent advancements in head-mounted displaying technologies, virtual reality pedestrian simulators have become a common tool for traffic safety research. In contrast to field studies, test track studies and traffic observations, simulators enable researchers to analyze pedestrian behavior in a safe and controlled environment. However, creating the necessary virtual environments is time-consuming, especially in terms of meeting today's expectations regarding graphical level of detail and realism. Furthermore, VR experiments often lack a body representation or require additional sensors to create an avatar. Due to the laboratory setting, VR simulators might fail to convey the feeling of standing on an actual street. In addition, simulators on the one hand and real-world testing on the other hand leave a methodological gap on the reality-virtuality continuum. This paper presents a novel approach for an augmented reality pedestrian simulator. With this simulator, the participant experiences virtual vehicles, augmented on a real scenario, allowing for safe and controlled testing in a realistic setting. In a between-subject design, 13 participants experienced a gap acceptance scenario with virtual vehicles, while 30 participants experienced the same scenario with real vehicles in the same environment. These participants were instructed to initiate a street crossing if they considered that the gap between the two experimental vehicles was safe to cross the street. Results indicate similar, but also offset behavior for both conditions. Lower acceptance rates and later crossing initiation times could be observed in the augmented reality condition. Still, it was shown that augmented reality renders a promising tool for pedestrian research but also features limitations depending on the use case.

*e-mail: philipp.maruhn@tum.de
†e-mail: andre.dietrich@tum.de
‡e-mail: lorenz.prasch@tum.de
§e-mail: sae.schneider@tum.de

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Mixed / augmented reality; Human-centered computing—Human computer interaction (HCI)—Empirical studies in HCI; Human-centered computing—Human computer interaction (HCI)—Interaction devices—Displays and imagers

1 INTRODUCTION

Traffic is becoming more and more complex. On one hand, increasing urbanization leads to a growing number of vehicles in already dense cities, where other road users are omnipresent. On the other hand, the advent of automated driving yields a new challenge for urban traffic, as non-motorized road users need to interact with cars driven by machines. Pedestrians are the most vulnerable group of road users. Although pedestrian fatalities in Germany decreased slightly by 5.2 % over the last year, they still account for 14.0 % of all traffic fatalities in 2018. Pedestrians involved in accidents are 3.5 times more likely to be fatally injured in comparison to vehicle drivers [25]. To ensure that a larger number of automated vehicles on urban roads do not increase pedestrian fatalities, but rather contribute to a reduction in accidents, traffic safety research is as important as ever.

Analyzing pedestrian behavior is a key aspect for understanding the cause of traffic accidents and predicting the effects of automated vehicles on urban traffic. Virtual environments are a commonly applied tool to study human behavior, especially within the context of traffic safety research. In contrast to field studies and observations, they allow for experiments in a safe and controlled environment. "Pedestrian simulators" or "street crossing simulators" enable researchers to analyze pedestrian behavior. Technological advances in Head-Mounted Displays (HMDs) allow an increasing number of researchers to study pedestrian behavior using virtual reality (VR).

This paper presents an approach for researching pedestrian behavior utilizing augmented reality (AR). A comparison between the newly developed AR simulator and a real-world experiment was made in a participant study.

2642-5254/20/$31.00 ©2020 IEEE
DOI 10.1109/VR46266.2020.00-53

2 State of the Art

To analyze individual aspects of pedestrian-vehicle encounters, controlled experiments are indispensable. In general, these experiments can be divided into real-world and simulator experiments.

Real-world experiments are either conducted on urban roads [23] or on test tracks [8, 18]. In both cases, the person controlling the vehicle is usually a confidant with the study, whereas the participant takes the role of a pedestrian. Real-world experiments require a large amount of effort, depend on actual vehicles and require certain equipment for accurately measuring kinematic values of the approaching vehicle as well as multiple experimenters. Furthermore, actual road crossings in real-world experiments could affect the study participants' safety. Therefore, a study participant's reaction to vehicle behavior (e.g. specific braking maneuvers) is usually measured by some type of indication, such as taking a step backward [18] or pressing a button [8], rather than actual crossings.

Other techniques have been used to analyze street crossings in real-world settings: in order to ensure a safe testing environment, Lee et al. [12] created a mock road parallel to the actual road equipped with two passing vehicles to analyze crossing behavior. They instructed the participants to cross the mock road as if they were crossing the actual road, using the vehicles on the actual road as indicators for when to cross the mock road. Schwebel, Gaines and Severson [24] implemented two protocols to analyze crossing behavior in a real and a semi-immersive virtual setting: the participants were either instructed to shout "Now!" or to take two steps toward the curb to signal their intent to cross the street.

Real-world experiments offer incomparable realism — study participants perceive an actual vehicle, on a real road with no artificial limitation of perception. However, reproducibility can be an issue in experiments, where the approaching vehicle is driven manually, as each driven trajectory will be slightly different. Furthermore, the generated metrics cannot be translated into actual road crossing behavior.

Simulator experiments offer a safe, reproducible and controllable way to research pedestrian behavior. This is particularly noteworthy when the behavior of specific, even more vulnerable populations such as children or the elderly is studied (e.g., [13]). In general, two types of pedestrian simulators using virtual reality can be distinguished: cave automatic virtual environments (CAVEs), projecting the virtual environment on screens surrounding the participants, and HMDs, visually decoupling participants from the real world.

A CAVE creates an immersive virtual environment using an arrangement of projection planes, usually in a cube shape (cf. [2, 10]). Thereby, size and number of projection planes vary as do the projection techniques (monoscopic or stereoscopic rendering). Combined with motion tracking systems, participants can interact with the environment via naturalistic walking, since the displayed perspective is updated by the tracked translation and rotation. Study participants experience the virtual environment while visually perceiving their own body, with no artificial obstructions to their natural field of view (FoV). Depending on the setup, CAVEs are often cost-intensive to purchase and maintain.

In recent years, head-mounted displays have undergone a leap in technology in terms of tracking accuracy, tracking space, FoV and resolution in combination with a significant price reduction. Available plugins allow for a simple integration of the HMDs with (game) engines to create virtual environments such as Unity3D, resulting in an increasing number of HMD-based pedestrian simulators (e.g. [4, 5]). Participants can experience an immersive virtual traffic scenario by wearing virtual reality glasses and travel by walking naturally. An often-mentioned limitation is the lack of a self-body representation in the virtual environment by means of an avatar. To account for that, HMD-based simulators have been combined with motion tracking [7] to virtually represent the body segments.

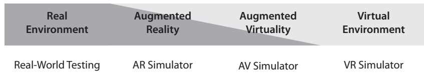

Figure 2: Methodical approaches for pedestrian behavior research categorized with regards to the reality-virtuality continuum (cf. [16])

Mallaro et al. [15] compared pedestrian street-crossing behavior between the two displaying technologies CAVE and HMD. The participants were instructed to cross a one-lane street with virtual vehicles passing at 23 mp/h. Results indicated that smaller gaps were accepted in the HMD setting (mean gap size 4.44 s) than in the CAVE (mean gap size 4.63 s).

Both technologies, CAVE and HMD, have their individual advantages but share one common drawback: in order to conduct studies, the virtual world must first be created. A higher quality level is associated with a higher degree of immersion and ultimately, higher presence scores [9], even though an increased level of detail has been reported to have a smaller impact on spatial presence than tracking level, stereoscopy and FoV [3]. Perez et al. [19] created an AR framework were pedestrians wearing a HoloLens are able to meet other user controlled and virtual agents in a virtual city. However, their main motivation for using the HoloLens as AR HMD was the freedom of movement. Still, the scenery was completely virtual. Depending on the scenery, creating the virtual environment and its adaptation to other traffic scenarios can be a time-consuming process.

In a typical street crossing experiment, the objects of interest are relatively limited. Often, only the moving vehicles are of importance, whereas the main purpose of other objects (houses etc.) is to give a proper setting in order to create a realistic backdrop. In contrast to VR, AR allows to make use of an existing scenery without the need to model the virtual surroundings. Apart from the reduced workload in creating virtual environments, it offers an option to evaluate scenarios in real-world locations. If, for example, the effect of a particular change in infrastructure should be evaluated, the physical location itself may be the testbed for this evaluation. Virtually replicating a given location, in contrast, bears the risk of impreciseness and the omission of vital objects.

Furthermore, both CAVE and HMD experiments are mainly conducted in laboratory rooms. The knowledge of being inside a closed room might break the immersion of the virtual scene and influence behavior (e.g. participants may show riskier behavior in street crossing tasks). Finally, AR allows to conduct studies in natural open-air environments. In particular for wireless systems, this allows researchers to cover larger areas without being restricted by the physical boundaries of laboratories.

Motivated by this, this work presents an approach to use AR for pedestrian simulators: instead of creating virtual scenery, virtual vehicles are superimposed on a real-world environment. If one assigns existing methods (real-world experiments and VR simulators) to the reality-virtuality continuum by Milgram et al. [16], one notices that currently only the extremes of this spectrum are covered. However, the role of AR in research has changed drastically in the past decade and technological advancements make it possible to apply AR in a wider range of use cases [11]. The presented AR approach in this work makes use of these new technologies and aims to bridge the methodological gap between real-world testing and VR simulator experiments (cf. figure 2).

Figure 3: Left: SRWorks SDK based on Vive Pro's built in cameras. Right: ZED Mini camera and SDK.

3 Apparatus

The aim of this work was to create an AR framework for traffic safety research from a pedestrian's perspective. Based on the previously established VR pedestrian simulators, the desired apparatus should meet similar requirements. The main purpose of the apparatus was to recreate a real-world experiment in AR: In the real experiment, two vehicles with a predefined gap (controlled via LiDAR) approach from a distance of about 320 meters and then pass the participant, who has to decide whether the gap is large enough for a road crossing. Various settings consisting of different hardware and software components were tested and evaluated for their applicability.

3.1 Hardware

Currently available AR headsets often lack a sufficient FoV, especially in the context of traffic safety research. Recent VR glasses, on the contrary, often feature a FoV of 110° up to 210° horizontally combined with reliable tracking based on reference stations (e.g. Valve Lighthouse Tracking). When connected to a stereo camera mounted on the HMD's front face, it is possible to display the surrounding, real environment via video pass-through. Video pass-through in the context of VR describes the process of live streaming the video signal from the HMD's (often built in) front-facing stereo cameras and render this video signal directly on the HMD's display. By fading the video pass-through, this technique is often used to warn users that they are too close to the boundaries of the predefined play area. Thus, this technique is a suitable method for transferring the user's attention from the virtual to the actual environment within a limited period of time and without putting the HMD off. Since the user only sees a video stream rendered on the VR glasses' display, one could argue that in this case it is more a matter of VR. However, video see-through can still be categorized as "See-through" AR displays [16].

The HTC Vive Pro features two front-facing cameras. First tests using the SRWorks SDK to access the cameras for AR video pass through seemed promising. However, the camera resolution of 480p made it difficult to identify objects at increasing distances. The virtual objects, however, are displayed in the HTC Vive Pro at full resolution (1440x1600 per eye). As a result, the virtual objects clearly stand out from the real environment, which is recorded and rendered at a lower resolution (cf. figure 3). To account for this, a Zed Mini stereo camera was mounted on top of the Vive Pro's integrated cameras to increase video pass through resolution to 720p. Combined with the Zed Mini's SDK, virtual objects are rendered at the same resolution as the camera's 720p to better blend virtual objects with the real environment. However, the Zed Mini's FoV (54° vertical, 85° horizontal) does not match the Vive Pro's FoV (110°). Since the Zed Mini's vertical FoV is noticeably smaller, users experience a black bar at the top and the bottom of the display (cf. figure 3). As a result, users have to bend their head forward even more than with the Vive Pro to look at their feet on the ground. The horizontal FoV, however, is not noticeably decreased. The trade-off between the Vive Pro's increased FoV and Zed Mini's higher image quality was decided in favor of the Zed Mini's higher resolution, since it seemed more important for the present use case to be able to identify real-world objects together with the virtual vehicles at greater distances. The software was hosted on a VR gaming PC running on an Intel(R) Core(TM) i7 8700k CPU with 32 GB Ram and a GeForce GTX 1080 Ti graphics card. The Zed Mini's manufacturer reports an expected latency of 30ms - 45ms at a 60Hz frame rate. As a result, the motion-to-photon latency was noticeable, primarily during fast and excessive head movements. However, movements of this kind were not expected during the experiment.

3.2 Software

The virtual environment and mixed reality framework were created with Unity 2018.2. Both tested SDK's, Vive SRWorks and ZED Mini, feature a plugin for a simple implementation in Unity. To facilitate a later calibration, the position of the virtual camera corresponds to the zero-point of the virtual environment. Laterally offset to this is the path the virtual vehicles are moving along. In more complex scenarios, this path could feature curves and acceleration or deceleration points. Changes in vehicle behavior are triggered as soon as the vehicle collides with such points. The path following is adapted from the Unity Standard Assets package. The wheels of the vehicles rotate based on the current speed.

The vehicle's behavior is based on a predefined set list. This set list was initially created for the real-world test track experiment, and the information gathered about speed, gap size for each trial and the trial order for all participants was stored. Based on this data, the virtual vehicles were spawned at the starting position, accelerated to the predefined speed and they opened the corresponding gap (whereas the gap in the real condition was controlled via a LiDAR range finder). The cars featured an engine sound combined with noise to simulate the wheel and wind noise.

During the trials, the vehicle's position as well as the HMD's translation and rotation, i.e. the movements of the participant, were stored in a buffer and periodically written to a CSV file. This data was then analyzed to extract the dependent variables (c.f. section 4.3), accepted crossings as well as crossing initiation time (CIT). The participants were also equipped with two Vive trackers mounted on top of their feet. These trackers record positional data based on the same technology as the HMD. Foot position was used to trigger the sound of the replicated virtual light barrier. In addition, positional data of the trackers was also recorded in order to calculate CITs based on foot instead of head movements. However, the tracking was not reliable enough to analyze the data. Due to the position of the reference stations to the sides of the participant, the trackers on the feet were covered by the participant's body. Therefore, the HMD's positional data was used to detect crossings and compute the CIT.

3.3 Calibration

The vehicles moved along a straight 320m-long path before passing the participant. In order to match the path of the virtual vehicles to the real street, the virtual environment needed to be calibrated as accurately as possible. The great length of the straight path in both directions only allowed for a few degrees of deviation. The trajectory of the vehicles along the path had to be matched within six degrees of freedom (position and rotation). Since the vehicles were bound to the path, only position and rotation of this one path had to be calibrated. Different approaches were tested for aligning the virtual with the real world for this framework.

3.3.1 Inside-Out Calibration

By using the information obtained from the stereo cameras, it is possible to extract the depth information and create a virtual mesh of the actual ground surface. By using wheel colliders and applying a gravitational force, it would then be possible for the virtual vehicles to drive on top of the surface. This would reduce the degrees of freedom to one remaining axis, since only the heading of the path remained to be adjusted. However, generating the virtual mesh based on camera information is always accompanied with artifacts, especially in outdoor conditions. Even by applying smoothing algorithms, it is difficult to control the outcome. However, future improvements in creating an environmental depth mesh based on stereo camera information would make this approach suitable. In addition, using the depth information provided by the camera for self-localization of the HMD would render further tracking hardware redundant.

3.3.2 One-Point Calibration

Since the inside-out tracking method, which is solely based on the stereo camera, was not reliable enough, Valve's lighthouse tracking was used. With this kind of tracking, the HMD orients itself based on the information provided by two (or up to four) light-emitting reference stations. This method requires an initial calibration: once the reference stations are set up, the HMD is placed at the desired zero-position. In this case, the HMD was placed on a level wooden board situated between the light barriers and placed next to the street. The light barriers were used in the "Real" condition to record crossing initiations. The HMD was oriented perpendicular to the street for calibration, which allowed the virtual vehicle's trajectory to the actual street to be matched. However, this method required an exact positioning of the HMD, resulting in an iterative and time-consuming process.

3.3.3 Mixed Calibration

In order to simplify the calibration process, a third option was tested. The headset was calibrated without being perfectly perpendicularly aligned with the street. To account for this inaccuracy, the virtual trajectory was adjusted afterward. While wearing the HMD with enabled video pass-through, the experimenter was able to orient the path with the help of a controller. The path's position and orientation were bound to the controller's. After placing the controller in the middle of the street, pointing toward the desired starting position, the experimenter was able to confirm the calibration by pressing a button. Unfortunately, the virtual cars were not easily identifiable at a large distances. Additionally, pointing the controller with the necessary accuracy was not possible. Therefore, the second option (calibrating the HMD as accurately as possible) seemed to be most promising and was applied for this setting.

It should be mentioned that all the above-described methods were manually executed. Another possibility would be to construct some sort of calibration device to either fix the headset (one-point calibration) or the controller (mixed calibration) to a predefined position. However, this possibility was rejected due to the conceptual character of the study and the lack of flexibility associated with such methods.

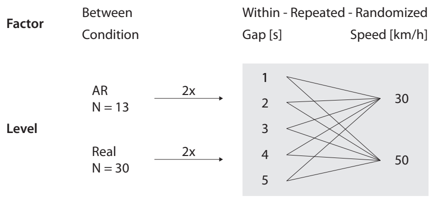

Figure 4: Study design: each subject underwent 20 trials (each combination of speed and gap twice) in a random order.

4 METHODOLOGY

The following study design (cf. figure 4) was repeated with two conditions, hereafter referred to as "Real" and "AR." In a between-subject design, initially 30 participants who were between 25 and 32 years old (mean age = 27.5, SD = 2.08, 15 females, 15 males) took part in the "Real" condition. Subsequently, data was collected from 13 other participants who were between 21 and 35 years old (mean age = 27.6, SD = 3.55, 6 females, 7 males) within the "AR" condition. The real-world data was collected as part of a larger research project to validate human behavior in different VR pedestrian-simulator settings. Due to the conceptual character of this work and organizational constraints, only 13 participants were tested within the "AR" condition. The study population for both conditions consisted mainly of university students or its employees. Data from both test parts ("Real" and "AR" condition) were recorded separately on the weekends in June 2019.

4.1 Environmental Setting

The tests were conducted in a large parking lot, located on the university campus. This area features a straight, paved street 450 m long and 4.8 m wide. The street was closed for the duration of the experiment, and since parking was also prohibited, no cars were parked along this street. Parking and driving were allowed in the neighboring lanes (cf. figure 5). Since barriers separated the active lanes from the experimental area, vehicles moving in the active lanes did not interfere with the experiment.

4.2 Study Design

In order to ensure as similar as possible settings for both the "Real" and the "AR" condition, the experimental procedure was duplicated wherever possible. After providing informed consent, the subjects underwent a vision test to ensure that they had sufficient vision. After successfully completing the visual test, each participant's walking speed was measured. Participants were instructed to cross two markings on the street, which were 4 m apart, six times at a leisurely gait and then six times at a rapid gait. Walking time

Figure 5: Experimental setting: subject standing next to the street, wearing the HMD with virtual vehicles approaching.

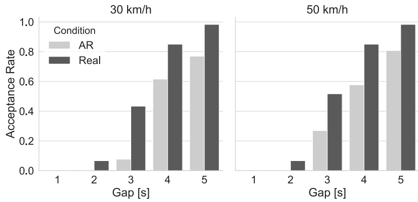

Figure 6: Gap acceptance rates for the “AR” and the “Real” condition for 30 and 50 km/h.

was measured with the same light barrier system. This was done for further data analysis within the real condition and repeated within the AR condition to ensure the same experimental procedure. Afterward, the participant was familiarized with the instructions: to initiate a street crossing when they felt safe crossing the street between the two passing vehicles. The cars started to accelerate and to adjust the experimental gap at a distance of 320 m. During the two practice trials, the vehicles passed the participant at 40 km/h and a 4 s gap and, during the second practice trial, with a 2 s gap. During the “Real” condition, the cars started to decelerate after passing the participant, turned and drove back to the starting position. In the “AR” condition, however, the vehicles disappeared after passing the participant at the end of the road and were reset at the starting position. As a result, the duration of the experiment was reduced by half (20 minutes in “AR” compared to 40 minutes in the “Real” condition). Time series effects during pedestrian gap acceptance experiments have to be expected with increasing experimental duration. However, it should be noted that the experiments were conducted in June, with daily maximum temperatures averaging 30° Celsius. Assuming that heavy sweating when wearing the HMD and the associated discomfort represent a further, possibly greater influencing factor, it was decided to reduce the experimental time described above.

After the two practice trials were completed, the actual experiment itself consisted of 20 trials. The two levels of speed (30 km/h and 50 km/h), combined with the 5 levels of gap (speed dependent-distance of 1, 2, 3, 4 and 5 seconds between the rear of the first and the front of the second car) resulted in 10 possible combinations of speed and gap. Each combination was repeated once. The resulting 20 trials were presented to the participant in a random order (cf. figure 4).

The University’s School of Medicine ethics committee approved this study.

4.3 Objective Measures

Subjects were instructed to start crossing the street after the first car had passed as long as they felt that they had enough time to do so before the second car got there. They were asked to ignore the fact that there were no other cars present before or afterward, and to only focus on the gap between the two oncoming vehicles. For safety reasons in the “Real” condition, subjects were instructed to only initiate a crossing by taking two to three steps toward the edge of the street and then turning around to avoid the chance of colliding with the cars. The instructions for the “AR” condition were given likewise.

The participant was positioned on the side of the street in front of a light barrier. The light barrier was triggered by the participants, when they initiated a street crossing. The light barriers were present but not functional in the “AR” condition. The VR glasses were calibrated with the zero position at the position of the physical light barrier.

During the trials, gap acceptance rates and CITs were measured as dependent variables. CIT was calculated by the time difference between the moment the rear of the first car passing the participant and the moment of the participant crossing the light barrier (or crossing the zero point in the “AR” condition).

4.4 Subjective Measures

To replicate the experimental procedure as closely as possible, the same post-trial questionnaires were used as in the “Real” condition. These include questions regarding the perceived safety of the potential street crossing, the subjective match of the crossing behavior in the experiment compared to the real-world behavior and questions about general remarks. In addition, items were included in order to subjectively evaluate the AR experience. These 5 items were answered directly after the experiment and putting off the glasses.

Presence questionnaires are commonly applied in VR experiments. However, due to the structure of their items, these questionnaires are not applicable for AR street crossing experiments. Regenbrecht et. al [22] created a 33 item questionnaire to measure mixed reality experiences. Out of this list, 5 relevant items had been extracted and adapted to suit the experimental design. At the end of the trials, the participants were asked to indicate their approval with each item on a 7 point likert scale, ranging from strongly disagree to strongly agree (cf. figure 9).

5 Results

A total of 860 gap acceptance trials were observed and analyzed, 600 in the “Real” condition and 260 in the “AR” condition. Effects of condition (“Real”, “AR”), gap size (1, 2, 3, 4 and 5 seconds) and speed (30 and 50 km/h) of the vehicles were analyzed on the dependent variables: crossing acceptance (dichotomous) and, when crossed, CIT (continuous in seconds).

5.1 Gap Acceptance Rates

In total, 366 potential crossings were observed — 285 crossings in the “Real” and 81 in the “AR” condition, which relates to overall gap acceptance rates of 47.5% for “Real” and 31.15% for “AR”. Gap acceptance rates are calculated by dividing the number of accepted gaps by the total number of gaps. The mean size of accepted gaps was higher in the “AR” condition (mean 4.40 s) compared to the “Real” condition (mean 4.16 s). Figure 6 illustrates the acceptance rates with regards to the gap sizes, grouped by the two levels of condition and separated by the two levels of speed. Descriptive analysis indicates a similar but offset distribution of acceptance rates with regards to gap size. A logistic mixed-effects model was created in R [21] to further analyze the effects of condition, gap sizes as well as the two different levels of speed. Individual participants were treated as random effect (random intercepts). The “Real” condition and the speed level of 30 km/h served as baseline. The model was computed according to equation 1 with the lme4 package [1].

Table 1: Results of the logistic mixed-effects model to predict the gap acceptance rates, computed according to equation 1.

	Estimate	Std. Error	z value	Pr(> $\|z\|$)
(Intercept)	-9.35	0.84	-11.09	0.00
ConditionAR	-2.52	0.72	-3.49	0.00
Gap	2.99	0.24	12.43	0.00
Speed.L	0.25	0.19	1.33	0.18

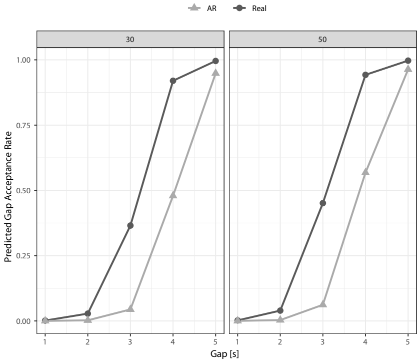

Figure 7: Predicted gap acceptance probabilities as outcome of a binomial logistical regression model created with ggeffects [14] in R [21].

$$CrossingProbability \sim Condition + Gap + Speed + (1|VP) \quad (1)$$

Results of the logistic mixed-effects model indicate a significant effect of gap size ($p < 0.001$) and condition ($p < 0.001$). The factor speed did not result in a significant effect (cf. table 1). A Durbin-Watson test [6] was conducted to test for auto-correlation of the residuals. The resulting value of 2.09 does not indicate auto-correlation. Figure 7 displays the predicted output of the model. The increase in acceptance rates in combination with larger gaps for the "Real" and "AR" conditions shows a similar trend. However, this trend seems shifted toward larger required gaps for equal acceptance rates within the "AR" condition as compared with the "Real" condition.

5.2 Crossing Initiation Time

CIT was calculated for the 366 observed crossings, as described in section 4.3. Figure 8 shows the distributions of CIT with respect to 30 km/h and 50 km/h, grouped by the two conditions "Real" and "AR." Descriptive analysis shows larger CITs for the "AR" condition (mean 30 km/h: 0.20 s, mean 50 km/h: 0.48 s) compared to the "Real" condition (mean 30 km/h: -0.36 s, mean 50 km/h: -0.17 s). In general, participants in the "AR" condition initiated their crossing after the vehicle completely passed, whereas the participants in the "Real" condition initiated their crossings before the first vehicle passed. In addition, speed showed a similar effect on both conditions, namely, higher velocity of the passing vehicle results in higher CITs.

A linear mixed-effects model was calculated in R [21] with the nlme package [20] to evaluate the effects of condition, speed and

Table 2: Results of the linear mixed-effects model to predict crossing initiation time (CIT).

	Value	SE	DF	t	p
(Intercept)	-0.33	0.10	321	-3.26	0.00
ConditionAR	0.65	0.14	40	4.54	0.00
Speed.L	0.14	0.02	321	6.55	0.00
Gap	0.02	0.02	321	0.88	0.38
ConditionAR:Speed.L	0.02	0.05	321	0.54	0.59

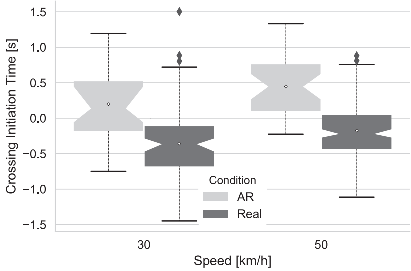

Figure 8: Crossing initiation times for the AR and the real world condition for 30 and 50 km/h. Mean values are represented by white diamonds, outliers by filled black diamonds. Notches indicate the 95% confidence interval of the median (bootstraped with 10000 iterations). The whiskers show the range of the data and extend up to 1.5 of the interquartile range.

gap size inference statistically. Table 2 summarizes the results of the model. Condition, speed and their interaction as well as gap size served as fixed factors ($CIT \sim Condition * Speed + Gap$). Participants were treated as random factor ($\sim 1|VP$). The model revealed a statistically significant effect for condition and speed. No significant effect was found for their interaction or for gap size (cf. table 2).

5.3 Subjective Measures

Figure 9 summarizes the results for each of the five mixed reality experience questionnaire items. In general, participants reported they rather felt part of a real than a virtual environment. The virtual vehicles seemed to be part of a virtual environment rather than the real world. The participants seemed to be inconclusive about the question if they share the same environment with the virtual vehicle. These items were answered by the 13 participants within the "AR" condition. In addition, all participants of both conditions filled out a post-trial questionnaire concerning the experience of the experimental setting.

Results of the questionnaire, asked in both conditions are summarized in figure 10. Only marginal differences between the two conditions were found regarding the perceived safety of a potential crossing. Participants rated an actual crossing similar safe for "Real" (mean = 3.43) and "AR" (mean = 3.23). Participants stated a similarly low probability regarding a collision with the passing vehicles in case of an actual crossing for "Real" (mean = 1.73) and "AR" (mean = 1.46). It is expected that the participants in the "AR" condition treated the virtual vehicles in the questionnaire the same way as real vehicles. When asked about the potential danger of a collision with one of the passing vehicles, again, only a small difference could be found between "Real" (mean = 2.96) and "AR"

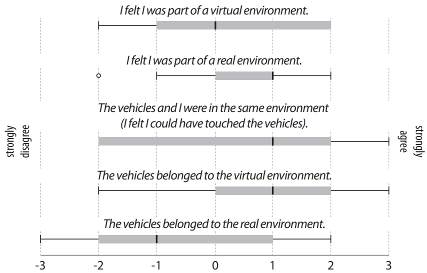

Figure 9: Distribution of responses (N = 13) to Mixed Reality Experience Questionnaire (MREQ) items.

(mean = 3.16, one missing value: N = 12) with the latter being even higher. By putting a cross on a 13 cm line without any markings, participants rated how similar they expected their behavior in real traffic conditions compared to the trials. Results showed a lower match for "AR" (mean = 8.78) than "Real" (mean = 9.90). The following item is missing the answers of two participants of the "AR" condition. Still, analogous results could be found when asked the same questions for trials they did not cross in the experiment (mean "Real" = 10.26; mean "AR" = 8.48, N = 11). Compared to normal traffic, participants rated their behavior slightly more risky for "Real" (mean = -0.13) than "AR" (mean = 0.15). Participants found it more difficult compared to normal traffic to decide to cross the street in "AR" (mean = 0.92) than in the "Real" condition (mean = 0.33). When asked about the necessary time to make a crossing decision, subjectively more time was needed in both conditions (mean "Real" = -0.43, mean "AR" = -0.62).

6 Discussion

This paper presents a novel approach analyzing pedestrian behavior in an AR setting. Overall, despite differences in behavior compared to the "Real" condition, the study demonstrated the applicability of the present setup. However, due to the deviation from the behavior in "Real" condition, results must be interpreted with caution. Further investigations are necessary to identify and quantify influencing factors. Still, it could be shown that AR marks a promising tool to investigate research questions concerning pedestrian behavior in a safe, controlled and above all realistic environment. In contrast to VR, it is no longer necessary to undergo the time and resource consuming process of creating the environmental setting. On the other hand, the framework requires a real environment. Depending on the research question or use case, this can drastically simplify or complicate the experimental setup. Furthermore, current technical limitations, which might be modified in future studies, are outlined in the following.

6.1 Technical Limitations

The virtual objects did not feature any reflections nor shadows (cf. figure 1). Both would be highly dependable on the actual real-world conditions with the potential to increase or break immersion. Where reflections of the sky could be easily realized with a virtual skybox, reflections of nearby objects (e.g. the participant) could only be created with a great deal of technical effort. Realistic shadows could be computed via a modeled position of a virtual sun, based on geospatial information (GPS location of the experiment) and the current time of day. This way, the virtual sun would correspond with the position of the real sun and enable the virtual objects to

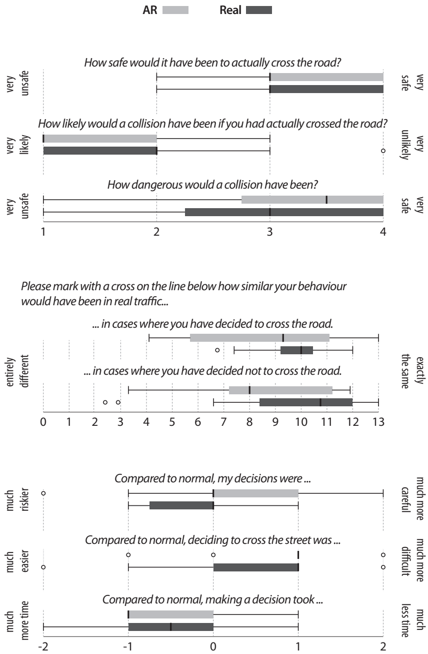

Figure 10: Distribution of responses subjective items, asked in both conditions.

cast shadows. However, this might require an even more accurate calibration to match the virtual shadows with the real environment.

The mounted stereo camera's resolution (ZED mini) of 720p in combination with the VR glasses' screen-door effect made it difficult for the participants to recognize the vehicles at great distances. The screen-door effect describes the visual artifacts (experiencing the virtual world like looking through a mesh) caused by the empty space between the pixels. The small distance to the displays in an HMD combined with the lenses increases this effect in VR headsets. Unstructured post-experimental interviews revealed that often the decision to cross the street was only made when the vehicles were closer than 150 meters.

The presented implementation did not feature mixed reality occlusion, i.e. virtual objects were not occluded by real objects in front of them. However, concerning this scenario, this denotes only a minor limitation, since the virtual vehicles were rarely behind real objects. In general, occluding virtual objects with real ones would be possible based on the depth information of the stereo camera. However, this technique cannot detect structures like hands reliably. In this case, combining the setup with an LED based hand tracker might improve occlusion.

The presented framework relied on wired connections of the HMD and of the stereo camera to the PC running the environment.

The initial setup without the additional camera used the HMD's built in cameras for video pass through. In this setting it would have been possible to replace the wired connection with the HTC's wireless adapter. However, the additional load due to the increased resolution of the added stereo camera does not allow for a wireless setup. Thus, during the experiment the HMD and the stereo camera were both connected with 5-meter cables to the PC. The cable length was sufficient for the instructed task, since the participants were only required to initiate a crossing. However, it cannot be completely ruled out that the tension of the cables has influenced the walking behavior.

Matching the virtual vehicles with the real environment represents the greatest challenge and at the same time the greatest point of criticism of this work. As described in section 3.3, calibration approaches were evaluated. However, in the end, the vehicles' trajectory did not fully match the real road, especially at greater distances. The virtual vehicles were precisely positioned on the road at the height of the test person. However, the vehicles' trajectory seemed to be radially distorted with increasing distance. As a result, the vehicles seemed to emerge from under the ground level, pass the participant at the correct elevation and then again submerge back into the ground as they move farther away from the participant. Further research must identify reasons for the radial distortion of the vehicle's trajectory. However, it is not to be expected that this error had a significant effect on the pedestrian's crossing decision. As reported, due to the technical constraints, it was only possible to identify the vehicles at smaller distances (approximately at less than 150m), by which point, the cars' position already matched the street level.

6.2 Overcoming Technical Limitations

VR experiments often lack a representation of the participant's body, i.e. an avatar. If, however, an avatar is present, additional body-worn equipment is often needed to track the body parts. In order to limit the required number of trackers, inverse kinematic models are applied to mirror the movements. These models, though, only approximate actual movement and may result in undesired movements of the avatar. In an AR framework, however, it is not necessary to create an virtual avatar because one's own body passes through the video feed. However, this presents new problems, such as occluding virtual objects with one's hands (cf. section 6.1).

Due to the conceptual state of this framework, there are further technical limitations, as described above. Most of these limitations can be addressed with further development of the presented framework. Reflections and realistic shadows can be added. Occlusion of virtual objects is potentially feasible with further technical equipment or better performing algorithms to analyze stereo camera's depth information. A further increase in resolution regarding the VR glasses displaying technology as well as stereo cameras for video pass through is to be expected. An increased resolution in combination with new displaying technologies would potentially reduce the screen door effect. This would allow vehicles to be recognized at greater distances. Currently, the FoV in AR headsets does not seem sufficient for analyzing pedestrian behavior. As long as this does not change, VR HMDs seem to be a suitable interim solution.

6.3 Validation

The selected use case for this proof of concept of an AR pedestrian simulator was a typical gap acceptance study. Since data was already recorded for a real-world test track scenario, this experimental design was replicated as closely as possible (see section 4.2). As a general remark, most of the participants found the experimental setting of only two cars passing by rather artificial, even though they were instructed to only consider the experimental gap and to disregard the missing traffic. Participants reported that the missing traffic before and after the experimental gap subjectively biased their crossing behavior. Although a setting with more cars is difficult to realize in the "Real" condition, future studies might employ alternative tasks and scenarios to evaluate the agreement between results obtained in the proposed AR simulator and in real world.

The responses to the Mixed Reality Experience Questionnaire vary widely for all items. Since this questionnaire is not validated yet, it is unclear whether this is a result of the questionnaire (misinterpretation of items) or the evaluated AR framework. Technical progress in displaying technologies will enable a growing number of scientists to use AR as research tool. Accordingly, better tools for their evaluation, in terms of validated questionnaires to rate AR experiences, can also be expected in the future.

In general, results of the objective measures indicate a similar but offset behavior for both conditions in terms of gap acceptance rates and CIT. The mean accepted gap sizes within the "AR" condition (mean 4.40 s) are in line with results by Mallaro et al. [15], especially compared to their HMD condition (mean 4.44). The created model (figure 7) predicts a similar increment in acceptance rates with increasing gap sizes with an offset of about one second. A similar picture emerges from the analysis of the CIT (figure 8). The effect of the two levels of speed could be found in both conditions. However, participants in the "AR" condition in general took longer to initiate a crossing. This effect is supported by the subjective rating in the post-trial questionnaire. The greater CITs in "AR" might be a result of the technical factors. Since the vehicles were only recognized at a certain distance, crossing decisions took longer. This is consistent with the fact that participants rated it more difficult to decide whether to cross compared to normal traffic in "AR" than in "Real."

7 Where To Go From Here

Using AR to investigate pedestrian behavior, there currently are a number of technical limitations to overcome. Not being able to identify vehicles in greater distances marked a main restriction in the current setting. Nonetheless, most research carried out in pedestrian simulators does not require such distances or velocities above 30 km/h. Currently, VR pedestrian simulators are commonly employed to evaluate new concepts to communicate from driverless vehicles to pedestrians, so called external HMIs (cf. [5,17]). The AR pedestrian simulator would allow to evaluate prototypes of external HMIs in a realistic setting without setting up physical prototypes. In general, AR seems a feasible tool in terms of a pedestrian simulator, which has the potential to bridge the methodological gap between real-world testing and VR experiments. With respect to the reality-virtuality continuum [16], it will be of interest to compare different technical settings (AR, VR, real world). Finally, this framework promises a wide range of applications for the research of human behavior, especially in the domain of human factors research (e.g. human-robot interaction).

Acknowledgments

This research was supported by the Deutsche Forschungsgemeinschaft (DFG). Project number 317326196

References

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

[2] V. Cavallo, A. Dommes, N.-T. Dang, and F. Vienne. A street-crossing simulator for studying and training pedestrians. *Transportation Research Part F: Traffic Psychology and Behaviour*, 61:217–228, 2019. doi: 10.1016/j.trf.2017.04.012

[3] J. J. Cummings and J. N. Bailenson. How immersive is enough? a meta-analysis of the effect of immersive technology on user presence. *Media Psychology*, 19(2):272–309, 2016. doi: 10.1080/15213269.2015.1015740

[4] S. Deb, D. W. Carruth, R. Sween, L. Strawderman, and T. M. Garrison. Efficacy of virtual reality in pedestrian safety research. *Applied Ergonomics*, 65:449 – 460, 2017. doi: 10.1016/j.apergo.2017.03.007

[5] A. Dietrich, J.-H. Willrodt, K. Wagner, and K. Bengler. Projection-based external human machine interfaces - enabling interaction between automated vehicles and pedestrians. In *Proceedings of the DSC 2018 Europe VR, Driving Simulation & Virtual Reality Conference & Exhibition*, pp. 43–50, 2018.

[6] J. Durbin and G. S. Watson. Testing for serial correlation in least squares regression. ii. *Biometrika*, 38(1/2):159–177, 1951.

[7] I. Feldstein, A. Dietrich, S. Milinkovic, and K. Bengler. A pedestrian simulator for urban crossing scenarios. *IFAC-PapersOnLine*, 49(19):239 – 244, 2016. 13th IFAC Symposium on Analysis, Design, and Evaluation ofHuman-Machine Systems HMS 2016. doi: 10.1016/j.ifacol.2016.10.531

[8] T. Fuest, L. Michalowski, L. Träris, H. Bellem, and K. Bengler. Using the driving behavior of an automated vehicle to communicate intentions - a wizard of oz study. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pp. 3596–3601, Nov 2018. doi: 10.1109/ITSC.2018.8569486

[9] J. Hvass, O. Larsen, K. Vendelbo, N. Nilsson, R. Nordahl, and S. Serafin. Visual realism and presence in a virtual reality game. In *2017 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, pp. 1–4, June 2017. doi: 10.1109/3DTV.2017.8280421

[10] Y. Jiang, E. E. O'neal, J. P. Yon, L. Franzen, P. Rahimian, J. M. Plumert, and J. K. Kearney. Acting together: Joint pedestrian road crossing in an immersive virtual environment. *ACM Trans. Appl. Percept.*, 15(2):8:1–8:13, Mar. 2018. doi: 10.1145/3147884

[11] K. Kim, M. Billinghurst, G. Bruder, H. B. Duh, and G. F. Welch. Revisiting trends in augmented reality research: A review of the 2nd decade of ismar (20082017). *IEEE Transactions on Visualization and Computer Graphics*, 24(11):2947–2962, Nov 2018. doi: 10.1109/TVCG.2018.2868591

[12] D. N. Lee, D. S. Young, and C. M. MCLaughlin. A roadside simulation of road crossing for children. *Ergonomics*, 27(12):1271–1281, 1984. doi: 10.1080/00140138408963608

[13] R. Lobjois and V. Cavallo. Age-related differences in street-crossing decisions: The effects of vehicle speed and time constraints on gap selection in an estimation task. *Accident Analysis & Prevention*, 39(5):934 – 943, 2007. doi: 10.1016/j.aap.2006.12.013

[14] D. Lüdecke. ggeffects: Tidy data frames of marginal effects from regression models. *Journal of Open Source Software*, 3(26):772, 2018. doi: 10.21105/joss.00772

[15] S. Mallaro, P. Rahimian, E. E. O'Neal, J. M. Plumert, and J. K. Kearney. A comparison of head-mounted displays vs. large-screen displays for an interactive pedestrian simulator. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, VRST '17, pp. 6:1–6:4. ACM, New York, NY, USA, 2017. doi: 10.1145/3139131.3139171

[16] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino. Augmented reality: A class of displays on the reality-virtuality continuum. In *Telemanipulator and telepresence technologies*, vol. 2351, pp. 282–292. International Society for Optics and Photonics, 1995.

[17] I. Othersen, A. S. Conti-Kufner, A. Dietrich, P. Maruhn, and K. Bengler. Designing for automated vehicle and pedestrian communication: Perspectives on ehmis from older and younger persons. In D. de Waard, K. Brookhuis, D. Coelho, S. Fairclough, D. Manzey, A. Naumann, L. Onnasch, S. Röttger, A. Toffetti, and R. Wiczorek, eds., *Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2018*, pp. 135–148, Oct. 2018. ISSN 2333-4959 (online). Available from http://hfes-europe.org.

[18] A. R. Palmeiro, S. van der Kint, L. Vissers, H. Farah, J. C. de Winter, and M. Hagenzieker. Interaction between pedestrians and automated vehicles: A wizard of oz experiment. *Transportation Research Part F: Traffic Psychology and Behaviour*, 58:1005 – 1020, 2018. doi: 10.1016/j.trf.2018.07.020

[19] D. Perez, M. Hasan, Y. Shen, and H. Yang. Ar-ped: A framework of augmented reality enabled pedestrian-in-the-loop simulation. *Simulation Modelling Practice and Theory*, 94:237 – 249, 2019. doi: 10.1016/j.simpat.2019.03.005

[20] J. Pinheiro, D. Bates, S. DebRoy, D. Sarkar, and R Core Team. *nlme: Linear and Nonlinear Mixed Effects Models*, 2018. R package version 3.1-137.

[21] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018.

[22] H. Regenbrecht, C. Botella, R. Baos, and T. Schubert. Mixed reality experience questionnaire (mreq) -reference. Technical Report Information Science Discussion Papers Series No. 2017/01, University of Otago, February 2017.

[23] F. Schneemann and I. Gohl. Analyzing driver-pedestrian interaction at crosswalks: A contribution to autonomous driving in urban environments. In *2016 IEEE Intelligent Vehicles Symposium (IV)*, pp. 38–43, June 2016. doi: 10.1109/IVS.2016.7535361

[24] D. C. Schwebel, J. Gaines, and J. Severson. Validation of virtual reality as a tool to understand and prevent child pedestrian injury. *Accident Analysis & Prevention*, 40(4):1394 – 1400, 2008. doi: 10.1016/j.aap.2008.03.005

[25] Statistisches Bundesamt (Destatis). Verkehr - Verkehrsunfälle 2018, 2019. Fachserie 8 Reihe 7. Online available from `https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Verkehrsunfaelle/Publikationen/Downloads-Verkehrsunfaelle/verkehrsunfaelle-jahr-2080700187004.pdf?__blob=publicationFile`.

frontiers in Virtual Reality

ORIGINAL RESEARCH
published: 20 December 2021
doi: 10.3389/frvir.2021.746971

VR Pedestrian Simulator Studies at Home: Comparing Google Cardboards to Simulators in the Lab and Reality

Philipp Maruhn *

Department of Mechanical Engineering, School of Engineering and Design, Chair of Ergonomics, Technical University of Munich, Munich, Germany

OPEN ACCESS

Edited by:
Ronan Boulic,
École Polytechnique Fédérale de Lausanne, Switzerland

Reviewed by:
Elizabeth O'Neal,
The University of Iowa, United States
Hyowon Lee,
Dublin City University, Ireland
Ilja T. Feldstein,
Harvard Medical School, United States

***Correspondence:**
Philipp Maruhn
philipp.maruhn@tum.de

Specialty section:
This article was submitted to Virtual Reality and Human Behaviour, a section of the journal Frontiers in Virtual Reality

Received: *25 July 2021*
Accepted: *11 October 2021*
Published: *20 December 2021*

Citation:
Maruhn P (2021) VR Pedestrian Simulator Studies at Home: Comparing Google Cardboards to Simulators in the Lab and Reality. Front. Virtual Real. 2:746971. doi: 10.3389/frvir.2021.746971

Virtual Reality is commonly applied as a tool for analyzing pedestrian behavior in a safe and controllable environment. Most such studies use high-end hardware such as Cave Automatic Virtual Environments (CAVEs), although, more recently, consumer-grade head-mounted displays have also been used to present these virtual environments. The aim of this study is first of all to evaluate the suitability of a Google Cardboard as low-cost alternative, and then to test subjects in their home environment. Testing in a remote setting would ultimately allow more diverse subject samples to be recruited, while also facilitating experiments in different regions, for example, investigations of cultural differences. A total of 60 subjects (30 female and 30 male) were provided with a Google Cardboard. Half of the sample performed the experiment in a laboratory at the university, the other half at home without an experimenter present. The participants were instructed to install a mobile application to their smartphones, which guided them through the experiment, contained all the necessary questionnaires, and presented the virtual environment in conjunction with the Cardboard. In the virtual environment, the participants stood at the edge of a straight road, on which two vehicles approached with gaps of 1–5 s and at speeds of either 30 or 50 *km/h*. Participants were asked to press a button to indicate whether they considered the gap large enough to be able to cross safely. Gap acceptance and the time between the first vehicle passing and the button being pressed were recorded and compared with data taken from other simulators and from a real-world setting on a test track. A Bayesian approach was used to analyze the data. Overall, the results were similar to those obtained with the other simulators. The differences between the two Cardboard test conditions were marginal, but equivalence could not be demonstrated with the evaluation method used. It is worth mentioning, however, that in the home setting with no experimenter present, significantly more data points had to be treated or excluded from the analysis.

Keywords: virtual reality, google cardboard, pedestrian simulator, road crossing, behavioral validity, pedestrians, bayesian statistics

1 INTRODUCTION

Pedestrian simulators are used in a similar way to driving simulators to explore pedestrian behavior in a safe and controlled environment. Participants experience a virtual traffic scenario from the perspective of a pedestrian. The hardware used to display these virtual worlds has undergone significant changes since the launch of pedestrian simulators.

Desktop-based applications have been used in a variety of studies to train children in pedestrian safety (McComas et al., 2002; Josman et al., 2008; Schwebel et al., 2008). The virtual environment is generally displayed either on a single screen or on three screens arranged in a circle. Due to the generally low hardware requirements, desktop solutions are often low-cost and flexible, but they also suffer from a low degree of immersion and limited possibilities of interaction.

CAVE-like (Cave Automatic Virtual Environments) systems, on the other hand, can overcome these limitations. Here, the participant is surrounded by projection screens on which a rear-projected virtual traffic scenario is displayed. The perspective of this virtual world changes as the participant moves his or her head. Possible interactions include natural walking, and crossing a virtual road, if permitted by the size of the structure (Mallaro et al., 2017; Cavallo et al., 2019; Kaleefathullah et al., 2020). A combination of stereo projectors and shutter glasses creates a 3D representation of the virtual scene (Mallaro et al., 2017; Kaleefathullah et al., 2020).

CAVE-like setups require a high implementation and maintenance effort and, they can be expensive, depending on the size and the hardware configuration. Head-mounted devices (HMDs) are available at a far lower price. Participants experience the virtual scenario trough VR glasses, and the displayed content is dynamically updated according to the position and rotation of the person's head. As long ago as 2003, Simpson et al., 2003 used an HMD to study the road-crossing behavior of children and young adults. They used the V8 by Virtual Research Systems, with a resolution of 640 × 480 per eye, a 60° diagonal field of view (FoV) and stereoscopic rendering (which was, however, not used in their experiment). Since then, the hardware has improved drastically. Since the release of consumer-grade VR glasses [cf. second wave of VR (Anthes et al., 2016)], HMDs have often been employed in pedestrian simulators, such as in the context of safety research (Deb et al., 2017) into smartphone distraction (Sobhani et al., 2017) or interaction with autonomous vehicles (Prattico et al., 2021), for instance.

An extensive overview of technologies and research designs used in pedestrian simulator studies in the last decade can be found in Schneider and Bengler (2020). All of these setups have in common that they are located in university laboratories or research facilities, which sometimes has implications for the demographic composition of the study sample. For instance, if certain groups are not explicitly addressed by the research question (e.g., traffic safety in the context of the elderly or children), recruitment often concentrates on the immediate environment, which results in an over-representation of healthy university students (Schneider and Bengler, 2020).

By providing low-cost alternatives, VR can be made accessible to a broader and more diverse set of subjects. Unlike conventional HMDs, mobile HMDs do not rely on an external computer (usually with high hardware requirements). Anthes et al. (2016) divide mobile HMDs into three categories: 1) standalone solutions that integrate the computing hardware into the headset and do not rely on any other technology, and devices that provide only a smartphone housing and use the phone's processing power and screen as an HMD. They further differentiate between 2) ergonomically designed cases and 3) simple cases, the latter generally offering lower degree of wearer comfort and a poorer optical display. However, the benefit of simple cases are the low acquisition costs. One prominent example of the latter variant are Google Cardboards, which were released in 2014. The combination of high smartphone ownership and low-cost Cardboards enables a wide range of applications, such as a training program to increase the safety of child pedestrians (Schwebel et al., 2017a; Schwebel et al., 2017b). The authors compared the Cardboard approach to a semi-immersive virtual environment with a sample of 68 college students (Schwebel et al., 2017a). The participants assessed both systems as having a similar degree of realism, and the Cardboard-based system was generally regarded by the authors as a usable and valid system.

In order to generalize results from simulator experiments to reality, a certain degree of validity is essential. Validity describes the degree to which observations in a simulator experiment match real-world behavior (Kaptein et al., 1996; Wynne et al., 2019; Schneider et al., 2021). Like driving simulators (Wynne et al., 2019), pedestrian simulators are in wide use, but validation studies are rare (Schneider and Bengler, 2020). There are two forms of validity (Wynne et al., 2019): absolute validity is when the same values are observed in the simulator and in reality, for example, for walking speeds. Relative validity is when the same effects are observed in both cases, even when the absolute values differ, for example, smartphone use influencing walking speed. Feldstein and Dyszak (2020) investigated decisions as to whether to cross a street in reality and in an HMD. Subjects stood at the edge of a single-lane road and a vehicle approached from the right. The subjects were asked to take a step backwards as soon as they judged the road to be unsafe to cross. The results could not confirm either relative or absolute validity, and in the virtual environment, smaller temporal distances were accepted. Unlike in the virtual environment, no effect in terms of different vehicle speeds was observed in reality. However, the relative validity of the effect of vehicle color (light vs dark) on the crossing decision was demonstrated in the very same study (Feldstein and Peli, 2020). Schneider et al. (2021) conducted a gap acceptance task on a test track, in a CAVE and in an HMD. By taking a step forward, participants signaled whether the gap between two vehicles was deemed safe enough to enable a single-lane road to be crossed. The most (i.e., also smaller) gaps were accepted on the test track. In both simulators, crossing was initiated later. Again, a correlation between increased vehicle speed and gap acceptance could be observed in both simulators, but not in reality, indicating that participants in the virtual environment relied on the total distance between the vehicles rather than on the temporal gap (Schneider et al., 2021). Similar results were observed in an augmented reality approach, with the same experiment (Maruhn et al., 2020).

Besides simulator properties, other effects can influence participant behavior and the transferability of the results to reality. Feldstein (2019) conducted a review of the technical, compositional and human factors at play when judging approaching objects in real and virtual environments. The presence of an experimenter or observer can also have an effect on the subjects' behavior. This phenomenon is known as the Hawthorne effect, for which a variety of definitions exist. For example, Oswald et al. (2014) describe the Hawthorne effect as "a change in the subject's normal behavior, attributed to the knowledge that their behavior is being watched or studied." For example, if subjects in a driving simulator feel observed, it can cause them to exhibit more socially desirable behavior (Knapper et al., 2015). One way of preventing this is to ensure complete anonymity. It has been shown in several psychological questionnaire studies that anonymously interviewed participants were more likely to report socially undesirable attributes (Lelkes et al., 2012). On the other hand, anonymity also reduces accountability and, in turn, the motivation to complete questionnaires accurately. Lelkes et al. (2012) compared three studies and confirmed that although, in some cases, anonymity led to an increase in socially undesirable responses, in all cases it led to lower accuracy and survey satisficing. The question is to what extent an anonymous setting influences the behavior displayed in a traffic simulation.

Motivated by the need for easier ways of testing diverse subject collectives and to validate the methods used on the basis of real data, this work replicates the study design by Maruhn et al. (2020) and Schneider et al. (2021) in a Google Cardboard setting, conducted under two different sets of conditions, in which one group of subjects does the experiment in a conventional laboratory session, while the other group performs the experiment at home with no experimenter present. By comparing results of this study with data from Maruhn et al. (2020) and Schneider et al. (2021), it is possible to assess the effects of a low-fidelity but cost-efficient hardware setup. The additional remote setting enables the influence of a laboratory setting with a human observer to be analyzed. This leads to the two research questions posed in this work: 1) How does a low-cost solution rank within the current pedestrian simulator hardware landscape and 2) What differences or similarities result from subjects performing the experiment alone at home as compared to a laboratory setting?

2 METHODS

The German-French PedSiVal research project involved cross-platform validation of pedestrian simulators. As is the case with many studies of pedestrian simulators (Schneider and Bengler, 2020), crossing decisions were also the subject of the present investigation. For this purpose, crossing decisions were recorded on a test track with real vehicles and compared with the results obtained using a CAVE, HMD (Schneider et al., 2021) and augmented reality (AR) (Maruhn et al., 2020). This study protocol is herein replicated with Google Cardboards (see **Section 2.1**) in two different environments: half of the subjects completed the experiment in a dedicated room located at the university in the presence of an experimenter (condition: CBLab). The other half completed the experiment at home, with no experimenter present (condition: CBRemote). The subjects were free to determine the timing of experiments conducted at home. Under the laboratory condition, on the other hand, a fixed appointment system was used. To prevent possible sources of error in the remote setting (e.g., ambiguities in the instructions or technical problems), the data were first collected under laboratory conditions.

2.1 Study Protocol

In the experiment, the subjects stood at the edge of a single-lane road at a distance of 0.65 m (cf. **Figure 1**). In each trial, two vehicles approached from the right at a constant speed. The actual experiment trials were preceded by two practice trials. The speeds varied between 40 km/h in the two practice trials and 30 km/h or 50 km/h in the actual experiment trials. In each individual trial, the gap between the vehicles was constant, but it varied from trial to trial from 1 to 5s. In the two practice trials, one 2s and one 4s gap was presented. Subsequently, every possible combination of the two speeds and five gap sizes was presented once in a random order, resulting in a total of 10 trials after the two practice trials. These combinations of vehicle gaps and speeds were thus identical to the data collected previously on the test track, with Cave, or with an HMD (Schneider et al., 2021), and AR (Maruhn et al., 2020). Likewise, the position of the participant and his or her distance to the road was approximately the same under all conditions (small variations were however possible since the participants in the other settings were able to move, whereas the position in the Cardboards was fixed).

The aim of this study is to evaluate a low-cost alternative to current, commonly applied approaches in pedestrian simulators such as consumer grade desktop-based VR HMDs and CAVEs as well as more recent approaches like AR. A simple Cardboard casing was used, without any padding. To limit discomfort while wearing the Cardboard, the duration of exposure to VR was minimized as far as possible. However, the aim was for the experiment to resemble the previous experiments as far as possible. Balancing these two objectives led to the following modifications compared to Maruhn et al. (2020) and Schneider et al. (2021): The number of trials was halved, and each combination of speed and gap size was presented once instead of twice. On the test track, the vehicles had to turn around after each run, drive back to the starting point and reposition themselves. However as in Maruhn et al. (2020), these waiting times were eliminated here. The vehicles disappeared at the end of the virtual road before being re-spawned at the starting point. Before starting their experiment, Maruhn et al. (2020) and Schneider and Bengler (2020) checked visual acuity using a simple paper-based test and determined each individual's walking speed for the subsequent purpose of calculating safety margins. However, this was omitted in the present study as it was

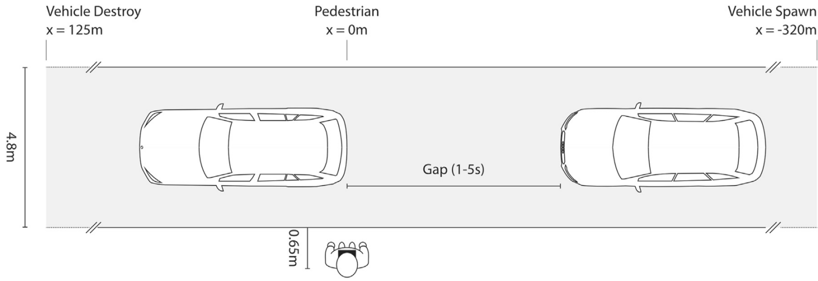

FIGURE 1 | Schematic sketch of the experimental setup. The virtual vehicles appear at a distance of 320 m on a 4.8 m wide road, pass by the subject at a predefined speed and distance, and disappear at the end of the road.

impractical in a remote setting in which there was no experimenter.

The overall experiment lasted about 30 min, including about 10 min of VR exposure. The study design was approved by the university's ethics committee.

2.2 Apparatus

Participants of both Cardboard conditions were provided with an unassembled Google Cardboard along with instructions and with a QR code for installing the mobile application. Participants were encouraged to install the application on their own smartphone. Participants under the lab condition who did not have their own Android smartphone were provided with a Nexus 5 device.

2.2.1 Cardboard Viewer

A Google Cardboard version 2.0 was used for this experiment (cf. **Figure 2A**). By pressing a button on the upper right of the Cardboard causes a pillowed hammer covered with conductive strip to be pressed against the smartphone display (Linowes and Schoen, 2016). This enables button presses can be registered as touch events. The Cardboard has two lenses (∅ 37.0 mm). It weighs 140*g*, has the dimensions (145 × 87 × 87 mm), and supports smartphone display sizes of between 3.7" and 6" (9.4–15.2 *cm*). The originally introduced Google Cardboards (2015) claimed to have a Field of View (FoV) of 80° and a lens diameter of ∅ 34.0 *mm* (Linowes and Schoen, 2016) in comparison to 85° in AR (Maruhn et al., 2020) and 110° in the HMD (Schneider et al., 2021). The Cardboard has a printed QR code, which the test persons were asked to scan before they started. This adapts the display to the lenses and dimension parameters of the device [cf. Linowes and Schoen (2016)]. Neither the distance to the display nor between the lenses are adjustable. The Cardboard can be worn hands-free using an adjustable elastic headband. In practice, however, the use of a headband is not recommended. Holding the Cardboard with the hands limits the head's rotation speed, which can prevent kinetosis (Linowes and Schoen, 2016). But as only a few rotations were to be expected in the present study design, mainly when the vehicles pass by, and as no interaction with the button on the Cardboard was necessary while waiting for the vehicles, the subjects were still provided with a headband. However, the participants were free to decide whether to use it (no instructions were provided for the headband) or to hold the Cardboard in place with their hands.

2.2.2 Mobile Application

The virtual environment from Maruhn et al. (2020) and Schneider et al. (2021) was transferred to an Android app (cf. **Figure 2B**), and the virtual environment was created in Unity 2019.3 and Google VR SDK 2.0. The experimental data was stored in an online database (Google Firebase). The app was distributed *via* Google Play Store.

On running the app, users stated whether they were working under the CBLab or the CBRemote conditions. They then had to confirm the subject information and informed consent. Users were asked to scan the QR code on the Cardboard. Since difficulties were observed when assembling the Cardboard under the lab condition, an explanatory stop motion animation was added for use under the remote condition. Finally, users entered their body height and the app switched to Cardboard mode. The virtual camera height was then adjusted to the subject's body height. To reduce the number of draw calls and thus increase the performance, the virtual parking lot environment was converted into a 360° image from this camera position. This meant that no translational movements were possible (however, only rotations can currently be reliably tracked in Cardboards anyway). The two vehicles remaining as the sole dynamic objects featured 3D spatial sound. Upon entering the virtual environment, text boxes were displayed to introduce participants to the experimental task.

2.3 Experimental Task

Participants were told that they could interact with the virtual buttons in the VR by pressing the physical button in the Cardboard and using a gaze pointer to select a virtual button. The gaze pointer is a small white circle in the middle of the screen, that is only visible while the instructions are being displayed. The subjects were informed that the two vehicles would always approach from the right, with a constant gap and speed.

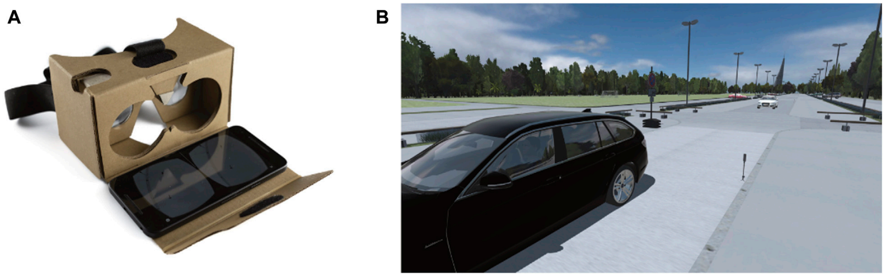

FIGURE 2 | (A) Google Cardboard Version 2.0 Viewer with a Google Nexis 5 Smartphone. **(B)** View from the subject's perspective during the experiment. Shown here is a 3-second gap at 30 *km/h*.

Participants were asked whether they were standing and using headphones for audio output. They were asked to rate only the gap between the two vehicles in terms of whether they considered it safe to cross the road. If so, they were asked to signal their intention to cross the street at the moment that they would commence crossing. Since it was not possible to cross the actual test track for safety reasons, only the intention to cross was assessed. In Maruhn et al. (2020) and Schneider et al. (2021), this was signaled by taking a step forward towards the street. Since it is not possible to reliably track translational movements with Cardboards, this was signaled by pressing a button at the upper right of the Cardboard in the present study. Each press of the button caused an audible beep sound to be emitted. Each button press was recorded in the online database to calculate the objective measurements.

2.4 Objective Measures

Crossing initiation time (CIT) and gap acceptance were recorded as objective dependent measures. Gap acceptance was recorded as a dichotomous measure, indicating whether or not participants deemed a gap large enough to cross. If a participant decided to cross, the CIT was also calculated. CIT describes the time difference from the moment the first vehicle has passed until the subject presses the button. A negative CIT thus means that a crossing was initiated before the preceding vehicle had completely passed by the subject. A button press such as the one shown in the situation in **Figure 1** would thus result in a CIT of 0s.

2.5 Subjective Questionnaires

After the trials were completed, the subjective data was recorded using the questionnaires contained in the app and results were also stored in the online database. The questions were the same as in Maruhn et al. (2020) and Schneider et al. (2021) and concerned demographics, how subjects rated the situations in which they did or did not cross the road, how easy they found it to make the decisions, and how likely it was that a collision would have occurred. The exact wordings can be found in **Figure 10**.

Online surveys or crowd-sourced data are often deemed to produce a low data quality as a result of careless behavior (Brühlmann et al., 2020). In long questionnaires, for example, attention checks are often used to identify deficient data sets. In this study, however, the focus was not only on the quality of the questionnaire but also on that of the simulator data. Another way of identifying careless behavior is to use self-reported single item (SRSI) indicators (Meade and Craig, 2012). At the end, participants were asked the following question, adapted from Meade and Craig (2012), using a continuous slider (extrema labels No–Yes): "It is critical for our study to include data only from individuals who give their full attention to this study. Otherwise, years of effort (by the researchers and other participants' time) could be wasted. Hand on heart, should we use your data for our analyses in this study?"

2.6 Recruitment

Members of the university (staff and students) were recruited as subjects. They were approached spontaneously on campus, largely without any personal connection. However, this led to considerable data collection problems in the remote setting. To ensure anonymity, no contact information was collected from the participants. They were only provided with the Cardboard and printed instructions on how to install the app. An email contact address was provided in case of any further questions. However, data from the first subjects recruited in this way were for the most part never received. Technical reasons can be excluded here, as continuous monitoring of the app did not indicate any problems. It appeared to be down to a lack of incentive or commitment resulting from the complete anonymity. To circumvent this problem, participants whose contact data was available from the extended social environment of employees and students were recruited for the remote component. One week after the Cardboards were given out, a reminder was sent to the subjects to perform the experiment at home. However, the experimental data remained anonymous. Subjects were only allowed to participate if they had not taken part in a previous study by Maruhn et al. (2020) and Schneider et al. (2021).

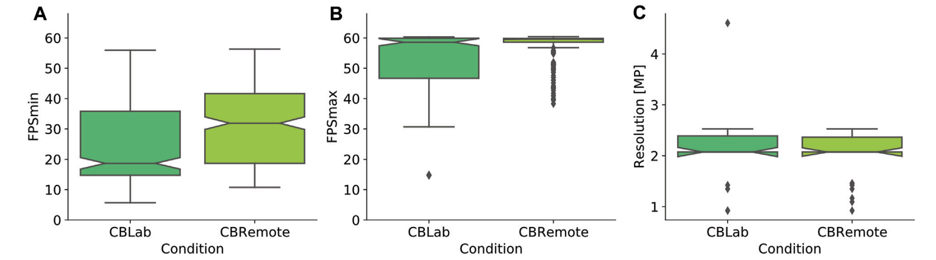

FIGURE 3 | Hardware performance indicators for the smartphones used in the two test environments. **(A)** shows the minimum frame rates that occurred in the test runs, and **(B)** the maximum frame rates achieved. **(C)** compares the display resolutions.

2.7 Study Population

Thirty subjects completed the experiment in each of the two test environments CBLab and CBRemote. Male and female participants were equally represented in both settings. The age distributions under the laboratory condition (M = 26.53, SD = 3.15, N = 30) and under the remote condition (M = 29.13, SD = 3.60, N = 30) were comparable to the data previously collected in reality (M = 27.50, SD = 2.08, N = 30), CAVE (M = 27.93, SD = 3.42, N = 30), HMD (M = 27.73, SD = 3.00, N = 30), and AR (M = 27.62, SD = 3.55, N = 13). Although the idea was to be able to recruit more diverse samples for the Cardboards used in a remote setting, this was not done here on grounds of comparability.

2.8 Hardware Performance

The current frame rate, measured in frames per second (fps), was continuously tracked during the trials. So as not to affect the performance, the frame rate was measured with a frequency of 4 Hz. The number of frames was divided by the elapsed time, i.e., the average frame rate of the previous 250 ms. The minimum and maximum frame rates were then logged for each trial. The phone's display resolution was also logged. The results are presented in **Figure 3**. Overall, the smartphones used under the remote condition were higher performing. The minimum frame rates measured (**Figure 3A**) are in many cases well below the frequently recommended level of 60fps (CBLab: M = 25.84*fps*, SD = 14.09*fps*, CBRemote: M = 31.85*fps*, SD = 13.29*fps*). However, it should be noted that these minimum frame rates usually only occur for a very short time, for instance, during a rapid head movement or when rendering complex 3D objects. Unfortunately, this generally happens when the vehicles are going past the test persons, who follow the vehicle movements with their heads. It can therefore not be completely ruled out that it affects the estimation task. In many smartphones, the maximum frame rate is limited to 60 fps (Linowes and Schoen, 2016), or even lower values, particularly in older models. The smartphones used under the remote condition achieved slightly higher maximum frame rates (M = 57.66*fps*, SD = 4.59*fps*) than those in the laboratory (M = 52.90*fps*, SD = 9.70*fps*), see **Figure 3B**. Regarding the display resolutions, only minimal differences between the smartphones used in the lab (M = 2.15 *MP*, SD = 0.62 *MP*) and those employed at home (M = 2.03 *MP*, SD = 0.47 *MP*) can be determined (**Figure 3C**).

2.9 Data Exclusion

If participants wished to mark a trial response as erroneous, for example, if they had clicked by mistake, they were asked to signal this by pressing the button three times. However, this did not happen in any of the trials. But there were several occasions on which a button was pressed well before or after the vehicles had passed. These cases were then deleted (CBLab: 2, CBRemote: 1). A CIT of ± 20s was defined as the threshold value. After this cleanup, a number of duplicate button presses still remained. In such cases (CBLab: 2, CBRemote: 8), the first of the two button presses was used in the onward analysis. Subsequently, there still remained cases in which the CIT was larger than the actual vehicle gap. Furthermore, very late button presses, i.e., those made shortly before the second vehicle passes cannot be considered as representing a time to start crossing. Consequently, CITs larger than half the gap size were excluded (CBLab: 2, CBRemote: 19) in CIT comparisons and only considered as gap acceptance. Under the laboratory condition, these data points were distributed among six subjects, and under the remote condition, among 13 subjects. It is possible that several of these excluded data points might occur for one subject. There was one participant in the remote group for whom five cases of double button presses occurred accompanied by five cases with excessively large CITs. This subject also accepted every gap. However, the cleaned data seemed plausible. Accordingly, it cannot be ruled out that the subject actually found all gaps passable and that the instructions had not been misunderstood. This subject was therefore not excluded from the analysis. The other cases were evenly distributed among the test subjects. Overall, more data points had to be cleaned under the remote condition than in the laboratory.

2.10 Data Analysis

Null hypothesis significance testing (NHST) is the predominant method of data evaluation used in human factors research, which includes driving simulator studies (Körber et al., 2016), although it does have some limitations and problems. NHST is often used to render a dichotomous statement, such as "Does an independent, manipulated variable have a significant effect on a dependent, observed variable?" This decision is then usually based on a previously defined α error level. The p value of a frequentist test indicates how likely the observation was, assuming the manipulated variable has no influence. However, the p value depends on a number of factors that must be taken into account by the researcher, such as sampling stopping rules and test selection (Kruschke, 2015a). Besides this single point estimate in the form of a p value, plausible limits around this estimator, i.e., confidence intervals, are also increasingly used in the social and behavioral sciences (Cumming and Finch, 2001). However, since confidence intervals are based on the same assumptions as p values, they suffer from the same limitations as those described above (Kruschke, 2015a). Moreover, confidence intervals only provide bounds, not a distribution function of the estimator (Kruschke, 2015a). This is in contrast with Bayesian statistics. Here, the outcome is actually a distribution function of the estimator. Bayesian approaches also avoid approximation assumptions (Kruschke, 2015b, p. 722) and allow for equivalence testing. These and other benefits have led to a rapid rise of Bayes methods in a variety of disciplines (Kruschke et al., 2012), but not in social and behavioral research [e.g. Kruschke et al. (2012); Körber et al. (2016)]. This work is a first step in applying Bayes statistics in the context of pedestrian simulator studies. Data analysis and model creation for this work were performed in Python, version 3.8.

3 RESULTS

As stated at the beginning of this work, this study replicates a design that has already been performed in reality, as well as in CAVE, HMD (Schneider et al., 2021), and AR (Maruhn et al., 2020). To better understand the results of this work in the overall context (Research Question 1), the data from previous studies are given in figures and analyses. The number of trials in the present setting was half that used in previous studies, and each combination of gap size and speed was presented only once. Accordingly, only the first 10 trials from the previous studies were considered. The data analysis focuses on two comparisons: the difference compared to the test track and the difference between the two Cardboard conditions (laboratory and remote, Research Question 2).

3.1 Gap Acceptance

In addition to the 300 crossings (10 trials x 30 participants) performed under each of the two Cardboard conditions (lab and remote), the analysis also includes crossings made in previous studies: REAL (300), CAVE (300), HMD (298, whereby 2 data points were excluded because the subjects were too far in front of the virtual light barriers) and AR (130, as there were only 13 subjects). **Figure 4** presents a summary of acceptance rates isolated for the variables: test environment (**Figure 4C**), gap size (**Figure 4D**), and speed (**Figure 4E**), as well as the combination of test environment and gaps (**Figures 4A,B**). In total, 654 of the 1,628 crossing opportunities were rated as passable, representing an acceptance rate of 40%. Acceptance rates for each factor were calculated from combinations of the remaining two factors (cf. **Figure 4**). Acceptance rates were highest in the real environment on the test track ($M = 0.48,\ SD = 0.5$) followed by CAVE ($M = 0.43,\ SD = 0.5$), CBRemote ($M = 0.42,\ SD = 0.49$), CBLab ($M = 0.39,\ SD = 0.49$), HMD ($M = 0.34,\ SD = 0.47$), and AR ($M = 0.28,\ SD = 0.45$). More gaps were accepted overall at 50 km/h ($M = 0.43,\ SD = 0.5$) than at 30 km/h ($M = 0.37,\ SD = 0.48$).

A mixed logistic regression model was generated to analyze the binary outcome of gap acceptance. The gap sizes were re-centered around 0 to accelerate data processing and simplify model interpretation. To account for this, the factor gap size is referred to as GapC in the model descriptions and summaries. The base model included condition (categorical, 6 levels: REAL, CAVE, HMD, AR, CBLab and CBRremote), speed (categorical, 2 levels: 30 and 50 km/h), gap size centered (continuous, 5 levels: −2, −1, 0, 1, and 2s) and participant as a random factor (ID). The REAL condition served as a baseline. Speed was treated as a categorical variable, since only two velocities were used in the experiment. Two additional models were created by adding random slopes for either gap size or speed as was one more model, in which two fold interactions of the condition with the speed and gap were added to the base model. The models were created with Bambi (Capretto et al., 2020) which is built on top of PyMC3 (Salvatier et al., 2016). Bambi allows for formula-based model specification and automatically generates weakly informative priors (Capretto et al., 2020). The Hamiltonian Monte Carlo algorithm in combination with No-U-Turn Sampler (Homan and Gelman, 2014) is then used to compute the posterior distribution depending on the priors and the likelihood function of the observed data. Each model was created with 4,000 draws and 4 parallel chains. The target accept (sampler step size) was increased to 0.99 to remove any divergences, especially in the more complex model with interactions. ArviZ (Kumar et al., 2019) was used to analyze and compare the Bayesian models created. The results are shown in **Table 1**. Despite being only ranked second, the base model will be used for further data analysis. Since there is only a marginal difference in the information criterion (loo: leave-one-out cross validation), the base model was chosen due to its simpler model complexity and parameter interpretation. The structure of the base model is illustrated in **Figure 5** as a distogram, according to the convention by Kruschke (2015b). The model's output is summarized in **Table 2**. No divergences occurred in the model. For none of the fixed effects did a rank normalized R-hat $\hat{R} > 1.00$ occur [cf. Vehtari et al. (2021)]. Visual inspection of the trace plots indicated a sufficient exploration of the parameter space and mixing of the chains.

Contrary to NHST, Bayes analysis provides a distribution of estimators instead of a point estimate. The further analyses are limited to the estimators for the different conditions. If the beta estimators for condition are considered in isolation, this corresponds to a model equation in which Gap x_S and Speed x_G are set equal to zero (cf. **Figure 5**). Since Gap was re-centered,

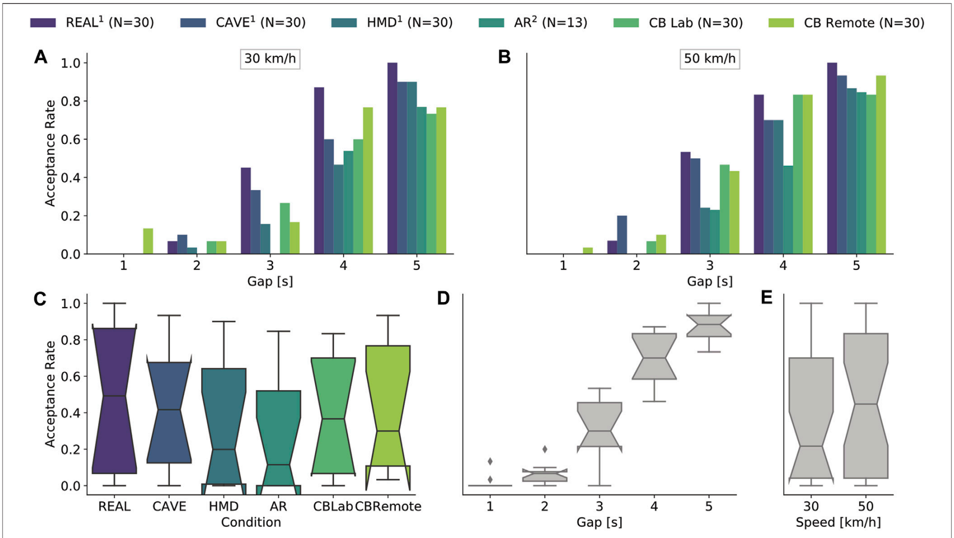

FIGURE 4 | Descriptive data of gap acceptance rates. **(A,B)** show the acceptance rates as a combination of vehicle gap and test environment subdivided for both levels of speed. The bottom three plots each represent one independent variable in isolation. The acceptance rates are shown as a function of **(C)** the experimental environment, **(D)** the gap sizes and **(E)** the levels of speed. The box plots show the distributions based on the other two respective factors. For example, **(D)** shows that there are combinations of test environment and speed that lead to a gap acceptance of 0 at 3*s*. In **(A)** it can be seen that these are AR at 30 *km/h*. The experimental data in REAL, CAVE and HMD were collected in Schneider et al. (2021). The experimental data in AR were collected in Maruhn et al. (2020).

TABLE 1 | Model comparison for predicting crossings as a function of condition, gap size, and speed as fixed factors and participant as a random factor. In addition to this base model, the other models include either a random slope for gap or speed, while one model features two-way interactions of gap and speed with condition.

	Rank	loo	p_loo	d_loo	Weight	se	dse	Warning	loo_scale
Random Slope Gap	0	−526.04	135.99	0.00	0.37	23.60	0.00	False	log
Base Model	1	−526.34	111.50	0.30	0.29	23.72	3.29	False	log
Random Slope Speed	2	−526.75	116.88	0.70	0.00	23.80	3.40	False	log
Interactions	3	−527.99	126.96	1.95	0.35	23.79	4.79	False	log

this corresponds to an estimate for 3s gaps at 30 *km/h* (baseline speed). The distributions of the beta estimators for the different conditions can be subtracted from each other. For example, if the CBLab distribution is subtracted from the CBRemote distribution, a conclusion can be drawn about how the estimation for a crossing differs between the two conditions relative to the REAL condition. The result of a linear combination from a logistic regression are given in logits. By applying an exponential function, the logits can be converted to odds. Dividing the odds by 1 + *odds* results in probabilities (**Equation 1**).

$$probability = \frac{e^{logit}}{1 + e^{logit}} \quad (1)$$

Of particular interest are the differences in the crossing probabilities between the simulators and reality (Research Question 1) as well as between the two Cardboard conditions (Research Question 2). **Figure 6** contrasts these differences in a ridge plot. The plot shows the respective posterior distributions' 94% highest density interval (HDI). HDIs are often given as 94% to emphasize their difference to confidence intervals. As with classical NHST, there are now several ways of describing the magnitude of these differences and, in contrast to NHST, about their equivalence as well (Kruschke, 2015b, p. 335 ff.). One way is to define a region of practical equivalence (ROPE). The boundaries of this region can be specified individually, depending on the use case. In this work, a data-driven approach was chosen. Kruschke (2018) defines ROPEs based

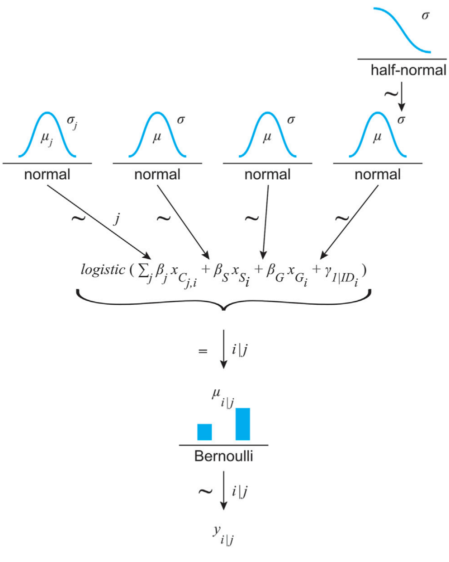

FIGURE 5 | Model structure for predicting each binary outcome of a possible crossing (y_i) depending on the simulation environment (j) and based on a Bernoulli distribution. Its μ value results from a linear logistic function with condition (C, 6 categorical levels referenced by index (j), speed (S) and gap size (G) as fixed effects. Due to the repeated measures design, the participants' ID was treated as a random effect (γ). The estimators were assigned weakly informative priors: normal distribution for the fixed effects and half-normal for the σ of the normal distribution for the random effect.

on the population standard deviation and half the size of a small effect ($\delta = \pm 0.1$) according to Cohen (1988). Since no standard deviation of the population is available in the present case, the sample standard deviation is used. With a dichotomous outcome variable, no standard deviation can be calculated initially. However, since each subject (i) repeated the experiment 10 times, the mean crossing probability of each subject ($\hat{p}_i$ = accepted crossings/number of repetitions) and the mean crossing probability of all subjects ($\hat{p}_0$) can be used to calculate the standard deviation of the crossing decisions of all participants (**Eq. 2**). The ROPE limits were then defined by multiplying $\sigma_{crossing}$ by $\pm$ 0.1 (**Eq. 3**). This results in a ROPE of ±0.018. Thus, differences of ±1.8% are treated as practically equivalent.

$$\sigma_{crossing} = \sqrt{\frac{\sum_{i=1}^{N} (\hat{p}_i - \hat{p}_0)^2}{N-1}} \tag{2}$$

$$ROPE = \pm 0.1 * \sigma_{crossing} \tag{3}$$

As can be seen in **Figure 6**, the combination of weakly informative priors and a relatively small set of data points results in very broad distributions relative to a narrowly defined ROPE. Accordingly, none of the distributions lies entirely within the ROPE and no practical equivalence can be assumed for any of the comparisons. The HDIs of HMD-REAL (*HDI* [− 0.426, −0.103]), AR-REAL (*HDI* [0.472, −0.149]) and CBLab-REAL (*HDI* [− 0.374, −0.026]) are completely outside the ROPE. Fewer gaps are accepted in these conditions than on the test track. For the remaining comparisons CAVE-REAL (*HDI* [−0.313, 0.069]), CBRemote-REAL (*HDI* [− 0.318, 0.059]) and CBRemote-CBLab (*HDI* [− 0.083, 0.213]), no statement can be made, because they are neither completely enclosed by the ROPE nor completely outside the ROPE. Makowski et al. (2019) describe another way of defining ROPEs for logistic models. For completeness, the evaluation is included in the analysis script provided online.

$$ROPE = \pm 0.1 * \frac{\pi}{\sqrt{3}} \tag{4}$$

They propose converting the parameters from log odds ratio to standardized difference using **Eq. 4**, resulting in a ROPE of ±0.181 on the logit scale. This approach yields the same results in this context, except that the CBLab-REAL comparison is just no longer completely outside the ROPE.

3.2 Crossing Initiation

Of the 654 gaps accepted, CITs were calculated for 633 deemed as potential crossings. Twenty-one cases were dismissed for which the CIT was greater than half the gap size (cf. **Section 2.9**). Since gap acceptance rates varied between the test conditions, so did the number of CITs: REAL (N = 145), CAVE (N = 128), HMD (N =

TABLE 2 | Summary for crossing model: *Crossing ~ Condition + Speed + GapC* +(1|*ID*). REAL served as a baseline. GapC refers to the centered gap size variable [–2s, 2s] rather than [1s, 5s].

	Mean	sd	hdi_3%	hdi_97%	mcse_mean	mcse_sd	ess_bulk	ess_tail	r_hat
Intercept	−0.579	0.362	−1.273	0.084	0.005	0.003	5966.0	9313.0	1.0
CAVE	−0.622	0.504	−1.563	0.333	0.006	0.004	7275.0	10392.0	1.0
HMD	−1.650	0.508	−2.606	−0.687	0.006	0.004	7025.0	9687.0	1.0
AR	−2.304	0.663	−3.502	−1.009	0.007	0.005	8201.0	10628.0	1.0
CBLab	−1.083	0.503	−2.013	−0.136	0.006	0.004	7559.0	10259.0	1.0
CBRemote	−0.666	0.501	−1.585	0.293	0.006	0.004	7071.0	10244.0	1.0
Speed	0.745	0.172	0.431	1.082	0.001	0.001	32756.0	10812.0	1.0
GapC	2.211	0.120	1.980	2.431	0.001	0.001	16983.0	13108.0	1.0

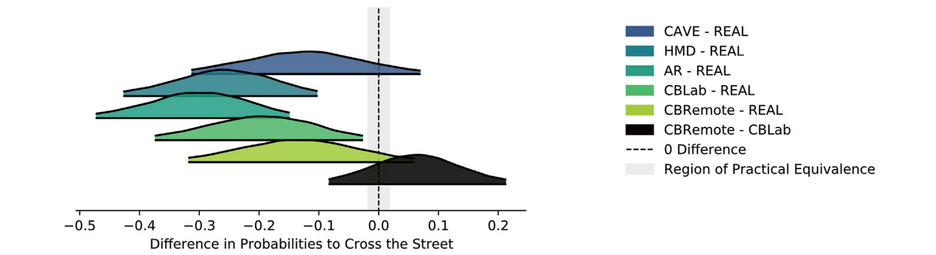

FIGURE 6 | Difference in posterior probabilities of crossing the street (for a 3s gap and at 30 *km/h*) between the simulators and the real-world setting, and between the two Cardboard settings.

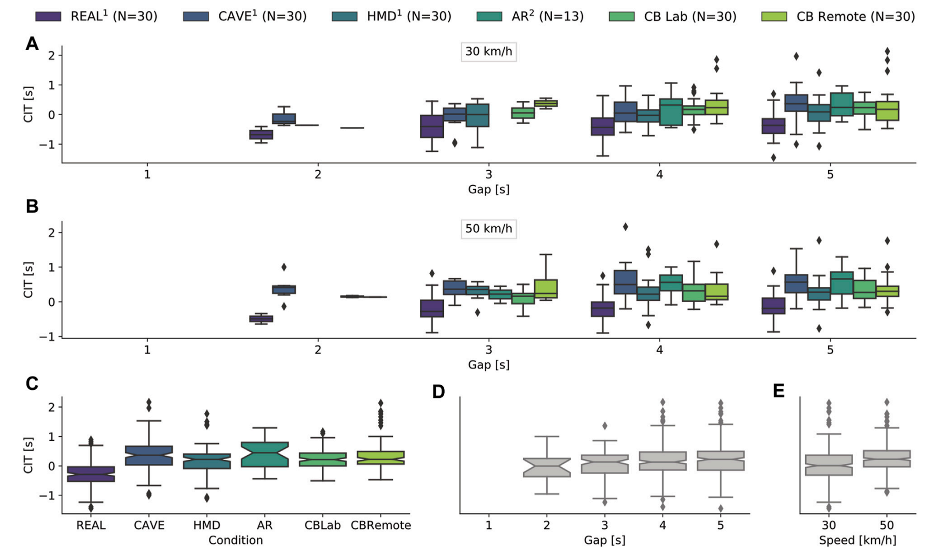

FIGURE 7 | Descriptive data of crossing initiation times (CIT). **(A)** and **(B)** show CITs as a combination of vehicle gap and test environment, subdivided for both levels of speed. The bottom three plots each represent a single independent variable. The CIT is shown as a function of **(C)** the experimental environment, **(D)** the gap sizes and **(E)** the speed levels. The experimental data in REAL, CAVE and HMD were collected in Schneider et al. (2021). The experimental data in AR were collected in Maruhn et al. (2020).

101), AR (N = 37), CBLab (N = 114) and CBRemote (N = 108). The overall mean CIT was 0.16s (SD = 0.52s), whereas in REAL (M = −0.27s, SD = 0.45s), distinctly lower CITs were observed than in the simulators: CAVE (M = 0.35s, SD = 0.51s), HMD (M = 0.16s, SD = 0.47s), AR (M = 0.38s, SD = 0.46s), CBLab (M = 0.24s, SD = 0.34s) and CBRemote (M = 0.34s, SD = 0.49s). As the gap size increased, CITs also increased slightly: 2s (M = −0.04s, SD = 0.48s), 3s (M = 0.06s, SD = 0.47s), 4s (M = 0.15s, SD = 0.52s) and 5s (M = 0.21s, SD = 0.52s). At 30 km/h, participants made the decision to cross earlier (M = 0.03s, SD = 0.55s) than at 50 km/h (M = 0.26s, SD = 0.46s). **Figure 7** shows the CITs as a function of condition (**Figure 4C**), gap (**Figure 4D**), speed (**Figure 4E**) or a combination of condition and gap for the two levels of speed (**Figure 4A** and **Figure 4B**).

In a similar way to the gap acceptance analysis above (cf. **Section 3.1**), different models were created with which to predict

TABLE 3 | Summary for CIT model: *CIT ~ Condition + Speed + GapC +* (1|*ID*). REAL served as a baseline. GapC refers to the centered gap size variable [−2s, 2s] rather than [1s, 5s].

	Mean	sd	hdi_3%	hdi_97%	mcse_mean	mcse_sd	ess_bulk	ess_tail	r_hat
Intercept	−0.430	0.073	−0.575	−0.301	0.002	0.001	2305.0	4860.0	1.0
CAVE	0.650	0.099	0.474	0.843	0.002	0.001	2416.0	5053.0	1.0
HMD	0.442	0.100	0.252	0.625	0.002	0.001	2457.0	5338.0	1.0
AR	0.638	0.136	0.374	0.885	0.002	0.002	3229.0	6233.0	1.0
CBLab	0.477	0.101	0.283	0.664	0.002	0.001	2397.0	4880.0	1.0
CBRemote	0.581	0.099	0.398	0.766	0.002	0.001	2763.0	5557.0	1.0
Speed	0.213	0.025	0.166	0.259	0.000	0.000	21679.0	12188.0	1.0
GapC	0.058	0.016	0.028	0.087	0.000	0.000	19890.0	13045.0	1.0
Sigma	0.300	0.010	0.282	0.318	0.000	0.000	13749.0	12077.0	1.0

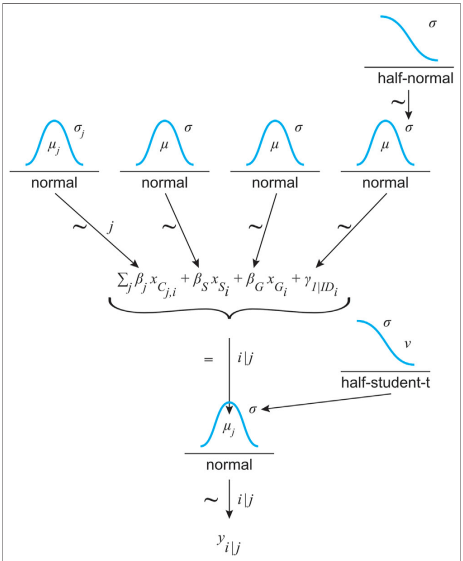

FIGURE 8 | Model structure for predicting CIT (y_i) depending on the simulation environment (j) and based on a Gaussian distribution. Its μ value results from a linear function with condition (C, 6 categorical levels referenced by index (j), speed (S) and gap size (G) as fixed effects. The standard deviation of the Gaussian distribution σ is estimated on the basis of a weakly informative normal prior distribution. Due to the repeated measures design, the participants' ID was treated as a random effect (γ). The estimators were assigned weakly informative priors: normal distribution for the fixed effects and half-normal for the σ of the normal distribution for the random effect.

CITs. Again, the base model featured condition, speed, gap size centered (GapC) plus the participant as a random factor (ID), while the additional models included either interactions or random slopes. The loo values for the models were again very similar, and therefore, the simple base model was chosen. The exact values can be found in the accompanying data analysis script. It should be noted that warnings appeared in the course of the model comparison [based on loo using Pareto-smoothed importance sampling (Vehtari et al., 2017)] to the effect that the estimated shape parameter of Pareto distribution was greater than 0.7 for some observations. Observations where $\kappa > 0.7$ indicate influential data points, and errors in loo estimation (Vehtari et al., 2017). Further analysis of these data points revealed that they are extreme values under the CAVE (N = 6), HMD (N = 6), and CBRemote (N = 3) conditions with comparatively very low or very high CITs. However, since the data points represent plausible CITs, they are not excluded from the model. The visual inspection of trace plots and auto-correlation plots, as well as the posterior predictive checks and the absence of divergences otherwise imply a sufficient model fit. The resulting model summary is presented in **Table 3** and the model schema is shown in **Figure 8**.

Here, too, further analysis focuses on the differences in the CITs between the simulation environments and reality and on comparing the two Cardboard conditions. The ROPE was determined on the basis of the standard deviation of all CITs as $\pm\, 0.1{*}\sigma_{CIT}$. Based on $\sigma_{CIT} = 0.52s$, this results in a ROPE of $\pm 0.052s$. **Figure 9** plots the difference in posterior distributions. Compared to the test track, the participants indicated their crossing decisions later in all simulated environments: CAVE-REAL (*HDI* [0.474, 0.843]), HMD-REAL (*HDI* [0.252, 0.625]), AR-REAL (*HDI* [0.374, 0.885]), CBLab-REAL (*HDI* [0.283, 0.664]), CBRemote-REAL (*HDI* [0.398, 0.766]). No statement can be made regarding a practical equivalence for the comparison between CBRemote and CBLab (*HDI* [− 0.090, 0.300]).

3.3 Subjective Data

Figure 10 presents all the questionnaire's results along with the wordings of the questions. **Table 4** gives an overview of means and standard deviations for each item and experimental setting. Overall, participants rated it quite safe to cross the street (Q1, Likert scale [1, 4]), with the highest scores achieved in reality on the test track, followed by AR, CBRemote, HMD, CAVE and CBLab. Collisions (Q2, Likert scale [1, 4]) were assessed as somewhat unlikely with the lowest values in AR, followed by CBRemote, REAL, CBLab, HMD and CAVE. The severity of collisions (Q3, Likert scale [1, 4]) was rated highest in AR followed by CBLab, REAL, CBRemote, HMD and CAVE. The two questions on similarity of behavior compared to reality (Q4 and Q5, continuous scale) were normalized to obtain values between 0 and 1. Both questions produced similar results. In

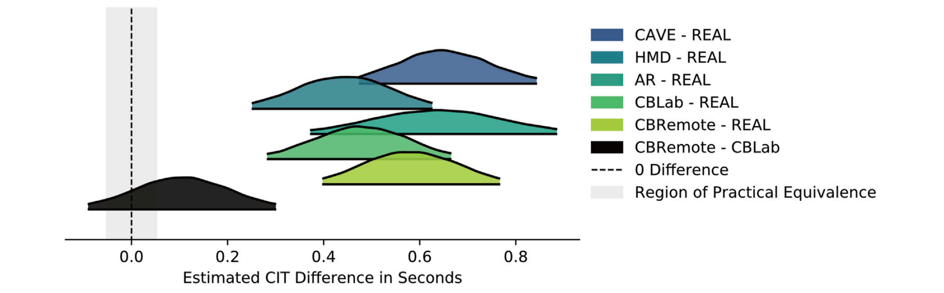

FIGURE 9 | Difference in posterior estimates for CIT (for a 3s gap and at 30 *km/h*) between the simulators and the real setting, and between the two Cardboard settings.

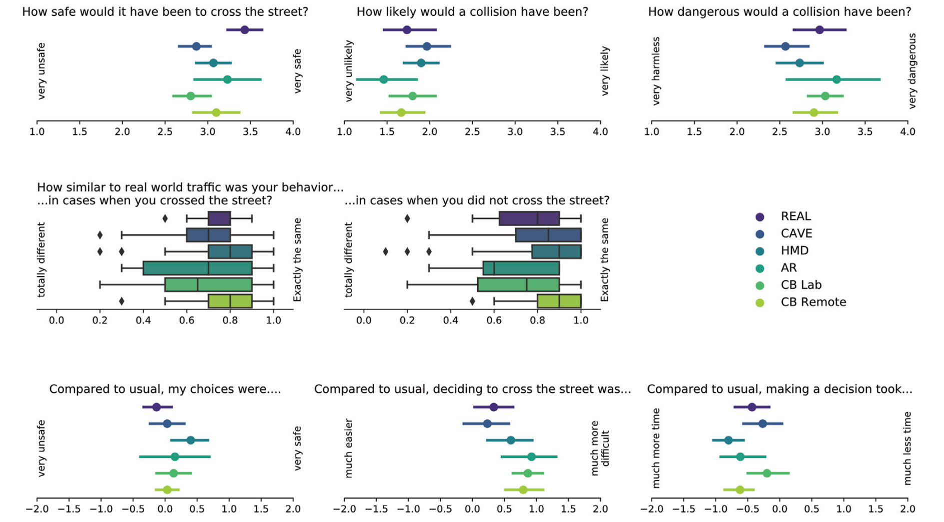

FIGURE 10 | Descriptive data from the questionnaires. Questions with Likert response scales are represented by point plots, questions with continuous response scales by box plots. For the Likert-scaled items, only the most extreme response options are indicated at the edges of the X axis in each case.

cases in which the participants crossed the street, behavior was rated the most similar in CBRemote followed by REAL, HMD, AR, CAVE, and CBLab. Again, in cases in which they did not cross, CBRemote scored highest, followed by CAVE, HMD, REAL, CBLab, and AR. This high correspondence to real road traffic situations was also reflected in the following question. The subjects indicated that their decisions were neither particularly safe nor unsafe compared to other situations (Q6, Likert scale [−2, 2]). Deciding to cross the street (Q7, Likert scale [−2,2]) was more difficult than usual, with the highest scores obtained in AR, followed by CBLab, CBRemote, HMD, REAL and CAVE. The participants also reported that taking this decision took longer (Q8, Likert scale [−2, 2]), being the shortest in CBLab, followed by CAVE, REAL, AR and CBRemote, and the longest in HMD.

Again, a Bayesian approach was chosen for analysis. For each question, a separate model was set up and fitted. In all models, the test environment served as the fixed (and only) factor, with REAL

TABLE 4 | Summary of questionnaire data. See **Figure 10** for questions wordings.

Condition	Q1		Q2		Q3		Q4		Q5		Q6		Q7		Q8	
	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std
AR	3.23	0.73	1.46	0.66	3.17	1.03	0.68	0.23	0.66	0.22	0.15	1.07	0.92	0.86	−0.62	0.65
CAVE	2.87	0.57	1.97	0.72	2.57	0.73	0.67	0.23	0.80	0.21	0.03	0.81	0.23	1.01	−0.27	0.87
CBLab	2.80	0.61	1.80	0.76	3.03	0.61	0.66	0.22	0.71	0.25	0.13	0.73	0.87	0.68	−0.20	0.92
CBRemote	3.10	0.76	1.67	0.76	2.90	0.71	0.78	0.17	0.87	0.16	0.03	0.50	0.79	0.82	−0.62	0.68
HMD	3.07	0.52	1.90	0.55	2.73	0.74	0.75	0.23	0.80	0.25	0.40	0.86	0.60	1.00	−0.80	0.66
REAL	3.43	0.57	1.73	0.83	2.97	0.89	0.76	0.09	0.77	0.22	-0.13	0.63	0.33	0.84	−0.43	0.73

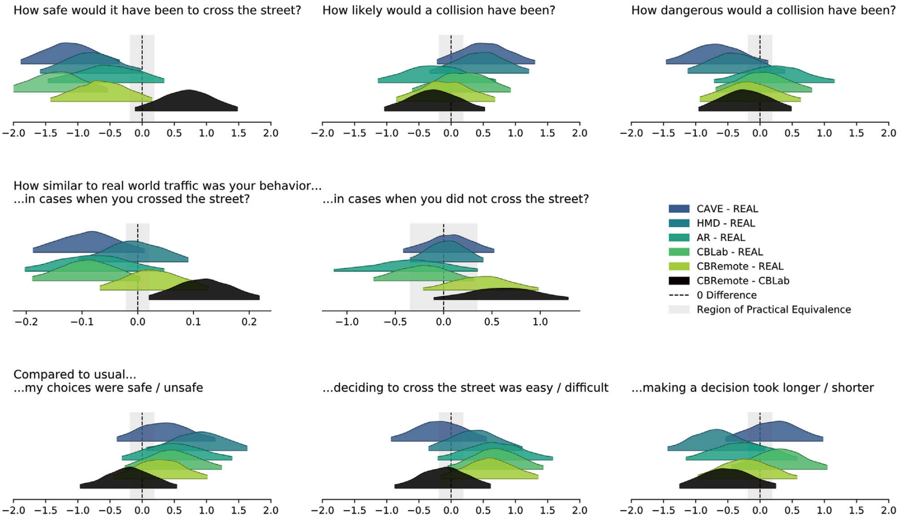

FIGURE 11 | Difference in posterior estimates for each of the subjective items in the questionnaire (cf. **Figure 10**). The results for Q1-3 and Q6-8 in the upper and lower rows are given in logits, the results of Q5 in the center plot on a log scale.

as the baseline. Likert-scaled items (Q1–3 and Q6–8, cf. **Figure 10**) were evaluated using ordered logistic regression. Since Bambi (Capretto et al., 2020) does not support ordered logistic regression at the time of writing, it was modeled in PYMC3 (Salvatier et al., 2016). Following McElreath (2020, Chapter 12) and consistent with previous analysis in this paper, weakly informative priors were used. Visual inspection of trace and auto-correlation plots indicated sufficient sampling, and posterior predictive plots evidenced an adequate fitting of the separate models. The two questions with continuous response scales were fitted with Bambi. Again, the test environment served as the sole factor. Here, posterior predictive checks indicated a poor fit between the Gaussian model and the heavily left-skewed data (cf. **Figure 10**). Data with a range from 0 to 1 was mirrored at 0.5 to obtain a right-skewed distribution and fitted to a Gamma distribution with log link functions to avoid divergences appearing with canonical link functions. Since these distributions do not support the observed values of 0, $1e-5$ was added to these data points. Ideally a truncated distribution that accounts for left-skewed data should be used. However, none is available at the current state of the libraries used. For both questions and the respective models, a model comparison revealed a higher model fit (based on loo values) for the Gamma models. However, the Gamma model for Q4 resulted in implausible posterior distributions under HMD and CBRemote conditions. The Gaussian model was therefore applied. Differences in posterior distributions were calculated to enable comparison of the two Cardboard conditions.

Similar to **Figures 6**, **9**, **11** shows the differences in estimator distributions for the above mentioned comparisons for each question. The outputs for the Likert-scaled items (Q1–3 and Q6–8) are displayed in logits, the results of Q4 on a log scale. In the case of logits, the ROPEs are defined by **Eq. 4**. The ROPEs' limits for the continuous items (Q4 and Q5) are again data driven and defined by the standard deviation of each response item: $\pm 0.1^{*}\sigma_{Question-Item}$ respectively $\pm 0.1^{*}\sigma_{\log(Question-Item)}$ to account for the Gamma model's log link function.

Crossings in CAVE (*HDI* [−1.886, −0.353]) and CBLab (*HDI* [−2.119, −0.539]) were rated unsafer than REAL based on a ROPE with the limits [−0.181, 0.181]. The results for HMD-REAL (*HDI* [−1.585, −0.013]), AR-REAL (*HDI* [−1.468, 0.340]), CBRemote-REAL (*HDI* [−1.419, 0.155]) and between the Cardboard conditions CBRemote-CBLab (*HDI* [−0.111, 1.481]) were inconclusive. Regarding ratings of how likely a collision would have been, neither differences nor equalities could be reported. The same applies to the questions relating to how dangerous a collision would have been. In cases in which the participants decided to cross, the behavior was rated closer to the real world traffic (*HDI* [0.020, 0.218]) under the remote Cardboard condition compared to the Cardboard under lab conditions based on ROPE with limits [−0.020, 0.020]. A similar trend could be observed in those cases when it was decided not to cross the street. The results regarding how safe or unsafe the choices were are also inconclusive, although participants tended to rate their crossings safer with the HMD than with REAL (*HDI* [0.066, 1.631]). Likewise, the results for the last two questions are inconclusive: i.e., 1) rating the decision as to whether to cross the street as easy/difficult and 2) how long it took to make the decision.

Finally, participants under the two Cardboard conditions were asked to rate the quality of their experimental data (SRSI, cf. **Section 2.5**) on a scale from 0 to 100. Under both conditions, the ratings were comparatively high: CBLab $M = 96.57$, $SD = 5.16$, CBRemote $M = 97.25$, $SD = 5.73$. The overall minimum was 74.88. Again, the distributions were left-skewed and thus mirrored and fitted with a Gamma distribution with log links (again to avoid divergences) and with the test environment as the only factor. The Gamma model performed better ($loo = 159.45$, $SE = 45.22$) showing a great improvement compared to the Gaussian base model ($loo = -178.24$, $SE = 10.79$). Again, $1e-5$ was added to values of 0 to allow the Gamma distribution to be fitted. The ROPE was again (cf. Q5) calculated based on the standard deviation of the observed, log transformed SRSI values. CBLab served as baseline in the model. The HDI for the effect of CBRemote compared to CBLab encompasses [−0.318, 0.281] and is thus completely within the ROPE ([−0.541, 0.541]).

4 DISCUSSION

This research was carried out to answer two research questions, namely, whether low-cost Cardboard headsets are a suitable substitute for high-end pedestrian simulator hardware and, secondly, whether they can be used to conduct studies in a remote setting with no experimenter present.

4.1 Cardboards and Other Simulators

The results observed in terms of gap acceptance were relatively similar to those in the other high-end simulation environments. Here, the two Cardboard conditions rank between the Vive Pro HMD and CAVE. Compared to the high-end HMD, the subjects accepted more gaps with the Cardboards and the results were thus more similar to those on the test track. However, it should be critically mentioned that the mode of signaling a crossing decision differs between Cardboards and other data. Since no translations could be tracked, detecting a step in the direction of the road was not possible, unlike in the other conditions and was instead signaled by pressing a button. Schwebel et al. (2017b) compared Cardboard button presses versus taking a step in a kiosk environment and detected a correlation in the number of missed crossing opportunities but only a correlation trend for CITs. Further studies should investigate the differences between button presses and naturalistic walking so as to render older HMD studies (mostly involving button presses) comparable with newer settings and their larger tracking areas. The need for such methodological comparisons can be clearly seen by comparing the results from these studies with those from Mallaro et al. (2017). In a similar research question, Mallaro et al. (2017) reported that the gaps accepted in an HMD were smaller than with a CAVE. The experimental task may also have had an influence here. In contrast to the study design presented in this work, in Mallaro et al. (2017), the street was crossed by physically walking. A comparison of accepted gaps is also interesting: while the overall rate of acceptance of a 3s gap was 34% here, 3s gaps were accepted much less frequently in Mallaro et al. (2017). The question arises what influence the range of the presented gaps have on the test person. Even if only smaller gaps are presented, the participant might feel compelled to accept gaps that he or she would actually consider too small just to fulfill the experimental task or expectations. However, another reason could be that participants receive feedback when they actually walk, so they might not cross the next time if they know that the time was not enough in a previous trial. It should be noted that the participants cannot see their own bodies in the Cardboards of any form. This is, however, possible in REAL, CAVE and AR as well as with a virtual avatar in HMD, by means of trackers positioned on their extremities, but this is not immediately feasible using a phone-based approach. However, the representation of one's own body seems to have an impact on the gap sizes accepted (Maruhn and Hurst, 2022).

Turning our attention to the CITs, the two Cardboard conditions fit in with the rest of the simulators. Particularly striking are the similarities of the two Cardboard distributions with the HMD distribution (cf. **Figure 9**). The Bayesian analysis confirmed that in all simulated environments, the crossing was initiated later than in the real-world setting on the test track. These results call into question the transfer of measured absolute values from simulation to reality [absolute validity, (Wynne et al., 2019)]. This is of particular relevance when determining safety-relevant measurements in road traffic, for example. Currently, this does not seem possible on the basis of virtual scenarios. The evaluation in Schneider et al. (2021) even questions relative validity: an influence of vehicle speed on gap acceptance could only be demonstrated in the two simulator environments (CAVE and HMD), but not on the test track (REAL). Not only the distorted distance perception in VR (Renner et al., 2013) but also the lack of resolution could still be a problem. Especially at large distances, the display of vehicles is reduced to a few pixels. Although this problem would be even more drastic under the Cardboard conditions, the differences in gap acceptance compared to the test track (**Figure 6**) and CIT (**Figure 9**) are similar to those in the other simulators. Other factors besides display resolution also seem to play a role.

Similarities were not only found with the objective data but also between Cardboards and the other simulators in the subjective ratings. In all simulators, participants indicated that it would have been less safe to cross the street than participants in REAL. This could explain why, overall, fewer crossings were accepted in the simulators and they were initiated later. The Bayesian analysis confirmed this difference for CAVE and CBLab. The fact that this was not confirmed for the remaining simulators is mainly due to the very broad distributions. There seems to be a trend that collisions were rated as being more likely to happen in CAVE and HMD but, at the same time, they would have been less dangerous. However, the results of the Bayes analysis do not allow for definitive differences. As for the first question, participants in all simulators rated their decisions as unsafer, even tough, this is only a trend. The participants seemed to have slightly greater problems making a decision with AR, HMD and the two Cardboard conditions, even though no definitive statements can be made. This can probably be seen as an effect of wearing an HMD.

4.2 Laboratory and Remote Setting

Approximately the same number of gaps were accepted under both Cardboard conditions with slightly more under the remote condition (cf. **Figure 6**), whereby the decisions were also taken slightly later (cf. **Figure 9**). However, the data quality from the remote setting must be viewed critically. More data points had to be treated than in the laboratory setting (cf. **Section 2.9**). The chosen criteria led to the exclusion of a large proportion of the implausible data, but some 1*s* gap acceptances from the remote setting remained (cf. **Figure 4A**). It was not certain whether a crossing was actually desired. For the evaluation of the CIT, however, these cases were excluded according to the defined rules. This is also one of the major disadvantages of a remote setting. Informal interviews after an experiment, which can help to check the plausibility of data, are no longer possible. Thus, it is not possible to determine in the remote condition whether the cause was misunderstood instructions, the uncontrolled setting, or, indeed, other effects.

Based on the Bayes analysis, no definitive conclusions can be drawn between the two Cardboard conditions regarding possible differences in objective measures. No differences occurred, but no practical equivalence could be demonstrated either. In particular, the combination of weakly informed priors, small amounts of data, and narrowly defined ROPEs meant that in none of the cases, neither for the objective nor for the subjective measures, does the ROPE completely enclose the distributions of differences between CBLab and CBRemote. Nevertheless, the results may be of value for future studies, for instance, for defining more informed priors or evaluating the data using other, practically dedicated, ROPEs.

While no data equivalences were demonstrated, there were some differences in the subjective measures. There seemed to be a trend that participants felt it was safer to cross the street in the remote setting. The differences between CBLab and CBRemote with the two continuous questionnaire items are worth highlighting. In both cases (when the participants crossed or did not cross the street), participants reported a high level of agreement with their everyday behavior, but this was even higher in the remote condition. Although no differences in objective data could be found, subjects in a home setting with no experimenter present seemed to subjectively perceive a higher degree of consistency with their everyday behavior. Even though more data points had to be treated in the remote setting (13 subjects vs 6 subjects in the lab), subjectively, participants rate their data quality as being equally high. Even if it was not subjectively perceived that way, the subjects may have performed the experiment with less accuracy or attention in the remote setting. It is also not possible to ensure whether participants in the remote setting were more distracted by external influences at home. Overall, however, it should be noted that the data in the two settings are very similar, and a remote setting can be considered comparable to a laboratory one and can even induce more subjectively realistic behavior. This again demonstrates how low cost, mobile-VR headsets can be seen as a suitable hardware device for experiencing virtual traffic scenes from a pedestrian perspective (Schwebel et al., 2017a; Schwebel et al., 2017b), even if it comes with some limitations.

5 LIMITATIONS AND IMPLICATIONS

This work marks a first attempt at conducting pedestrian simulator studies in a remote setting. Naturally, there are also limitations. The young age group across all experiments does not allow a generalization of the results to children or older adults. However, limitation to one age group was necessary to control for inter-individual differences in this between-subjects design (Schneider et al., 2021). A within-subject design would have increased the statistical power, but this was not possible and could have induced new effects (learning effects, comparative judgments) and it can limit comparison with previously collected data. Many optimizations were made to achieve a sufficiently high frame rate on mobile devices, but this could not always be ensured. However, it can be assumed that smartphone performance and display resolution will continue to increase over the next few years and that even complex virtual environments can be displayed smoothly. In the future, additional cell phone sensors could enable reliable tracking of translational movements in addition to head rotation, enabling other forms of interaction than button presses. Some participants took the opportunity to provide feedback on the experiment in a text input field in the questionnaire. Five users said that the vehicles started off at too great a distance or that there was too much waiting time between trials. This was also the case in other environments, where the waiting time between trials was even longer, but the vehicles were discernible at greater distances. In contrast, the lower display resolution of some of the smartphones meant that vehicles were not even recognizable at their starting position. In order to minimize the discrepancies with the other studies, these waiting times were nevertheless included. Three participants stated that the image was slightly blurry. This could be due either to low screen resolution or to the presence of screen protectors. Another limitation in imaging is that the lens distance could not be adjusted to the individual interpupillary distance (IPD). In the case of a significant deviation (for example in a child with a very small IPD) this can lead to visual discomfort (Peli, 1999) and it can also potentially affect depth perception (Woldegiorgis et al., 2018; Hibbard et al., 2020). However, in this study, none of the participants reported experiencing artifacts such as double vision. Future studies should evaluate the degree to which distance perception is influenced, whether this can be countered by modifying the virtual image to adjust for stereopsis, the necessity of excluding subjects displaying significant deviations, and the use of Cardboards with adjustable lenses. It should be noted that the use of the test subjects' own smartphones can itself lead to significant differences. For example, the instructions told participants to set the display brightness to the maximum level, but this was not ensured in any way. Two participants explicitly stated that it was difficult to estimate distance and time in VR. However, this was also stated by two test subjects using the high-end headset (under HMD conditions). In contrast to Schwebel et al. (2017b), two of the Cardboard participants reported simulator sickness symptoms. Two subjects also reported suffering from sensor drift, i.e., the environment continued to rotate slowly even without any head rotation. These two problems seem to be directly linked. In future studies, the sensors should be calibrated before commencing the experiment to minimize this problem.

Most of these limitations, also encountered with the other simulators, are either technical in nature and will potentially be solvable with advances in smartphone technology, else they are due to the nature of the experiment. Using Cardboards in a remote setting appears to be a feasible method of collecting data from otherwise underrepresented study populations (Schneider and Bengler, 2020) and it is also suitable for gathering larger sample sizes. However, it must be ensured that the participants are be able to carry out the experiment independently. This can be done to a certain extent during the development of the experiment, but for specific groups of people who are required to have support, a classic laboratory setting with an experimenter still seems to be more suitable. It has also been shown that the method of subject recruitment is crucial for the success of the data collection. Complete anonymity seems to elicit too little commitment, which, as in this case, can lead to subjects simply not completing the study. Even subjects with whom there was social contact needed continuous reminders to perform the experiment on their own. In this respect, the remote setting is very different from a trial in a laboratory, in which the timing is determined by a fixed appointment. For future trials in remote settings, I would therefore suggest a procedure that combines the advantages of remote and laboratory settings. For example, an appointment can be made directly with anonymously recruited subjects when the trial is conducted at home, or reminders can be sent from the app for this purpose. Furthermore, the recruitment of subjects via social contacts may well have favored the high SRSI values. Considering mechanisms such as social desirability, a completely anonymous setting could well lead to lower SRSI values.

Inevitably, more effort has to be put into the development of a remote test, since no experimenter is available to help and the instructions have to be unambiguous and self-explanatory. However, this yields the benefits of consistent instructions and a standardized test procedure.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: OSF | PedSiVal Mobile: https://osf.io/47nc9/?view_only=b4e5fb884abc43129a911142625cf846.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee at the Technical University of Munich (TUM). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

PM: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing—original draft, Writing—review and editing.

FUNDING

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)–Projektnummer 317326196. The publication fees were covered by the Technical University of Munich (TUM) in the framework of the Open Access Publishing Program.

ACKNOWLEDGMENTS

I would like to thank Stephan Haug (TUM|Stat) for advice on the statistical analyses and Sonja Schneider for providing valuable input on conceptualizing the study and for reviewing the manuscript.

REFERENCES

Anthes, C., Garcia-Hernandez, R. J., Wiedemann, M., and Kranzlmuller, D. (2016). "State of the Art of Virtual Reality Technology," in 2016 IEEE Aerospace Conference, 1–19. doi:10.1109/AERO.2016.7500674

Brühlmann, F., Petralito, S., Aeschbach, L. F., and Opwis, K. (2020). The Quality of Data Collected Online: An Investigation of Careless Responding in a Crowdsourced Sample. *Methods Psychol.* 2, 100022. doi:10.1016/j.metip.2020.100022

Capretto, T., Piho, C., Kumar, R., Westfall, J., Yarkoni, T., and Martin, O. A. (2020). Bambi: A Simple Interface for Fitting Bayesian Linear Models in python. Available at: https://bambinos.github.io/bambi/main/index.html#citation.

Cavallo, V., Dommes, A., Dang, N.-T., and Vienne, F. (2019). A Street-Crossing Simulator for Studying and Training Pedestrians, Special TRF issue: Driving simulation. *Transportation Res. F: Traffic Psychol. Behav.* 61, 217–228. doi:10.1016/j.trf.2017.04.012

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates.

Cumming, G., and Finch, S. (2001). A Primer on the Understanding, Use, and Calculation of Confidence Intervals that Are Based on central and Noncentral Distributions. *Educ. Psychol. Meas.* 61, 532–574. doi:10.1177/0013164401614002

Deb, S., Carruth, D. W., Sween, R., Strawderman, L., and Garrison, T. M. (2017). Efficacy of Virtual Reality in Pedestrian Safety Research. *Appl. Ergon.* 65, 449–460. doi:10.1016/j.apergo.2017.03.007

Feldstein, I. T., and Dyszak, G. N. (2020). Road Crossing Decisions in Real and Virtual Environments: A Comparative Study on Simulator Validity. *Accid. Anal. Prev.* 137, 105356. doi:10.1016/j.aap.2019.105356

Feldstein, I. T. (2019). Impending Collision Judgment from an Egocentric Perspective in Real and Virtual Environments: A Review. *Perception* 48, 769–795. doi:10.1177/0301006619861892

Feldstein, I. T., and Peli, E. (2020). Pedestrians Accept Shorter Distances to Light Vehicles Than Dark Ones when Crossing the Street. *Perception* 49, 558–566. doi:10.1177/0301006620914789

Hibbard, P. B., van Dam, L. C. J., and Scarfe, P. (2020). "The Implications of Interpupillary Distance Variability for Virtual Reality," in 2020 International Conference on 3D Immersion (IC3D), 1–7. doi:10.1109/IC3D51119.2020.9376369

Homan, M. D., and Gelman, A. (2014). The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* 15 (1), 1593–1623. doi:10.5555/2627435.2638586

Josman, N., Ben-Chaim, H. M., Friedrich, S., and Weiss, P. L. (2008). Effectiveness of Virtual Reality for Teaching Street-Crossing Skills to Children and Adolescents with Autism. *Int. J. Disabil. Hum. Develop.* 7, 49–56. doi:10.1515/ijdhd.2008.7.1.49

Kaleefathullah, A. A., Merat, N., Lee, Y. M., Eisma, Y. B., Madigan, R., Garcia, J., et al. (2020). External Human-Machine Interfaces Can Be Misleading: An Examination of Trust Development and Misuse in a CAVE-Based Pedestrian Simulation Environment. *Hum. Factors* 0, 18720820970751. doi:10.1177/0018720820970751

Kaptein, N. A., Theeuwes, J., and van der Horst, R. (1996). Driving Simulator Validity: Some Considerations. *Transportation Res. Rec.* 1550, 30–36. doi:10.1177/0361198196155000105

Knapper, A., Christoph, M., Hagenzieker, M., and Brookhuis, K. (2015). Comparing a Driving Simulator to the Real Road Regarding Distracted Driving Speed. *Eur. J. Transport Infrastructure Res.* 15, 205–225. doi:10.18757/EJTIR.2015.15.2.3069

Körber, M., Radlmayr, J., and Bengler, K. (2016). Bayesian Highest Density Intervals of Take-Over Times for Highly Automated Driving in Different Traffic Densities. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* 60, 2009–2013. doi:10.1177/1541931213601457

Kruschke, J. K., Aguinis, H., and Joo, H. (2012). The Time Has Come. *Organizational Res. Methods* 15, 722–752. doi:10.1177/1094428112457829

Kruschke, J. K. (2015a). "Null Hypothesis Significance Testing," in *Doing Bayesian Data Analysis*. Editor J. K. Kruschke. Second EditionSecond edition edn (Boston: Academic Press), 297–333. doi:10.1016/B978-0-12-405888-0.00011-8

Kruschke, J. K. (2018). Rejecting or Accepting Parameter Values in Bayesian Estimation. *Adv. Methods Practices Psychol. Sci.* 1, 270–280. doi:10.1177/2515245918771304

Kruschke, J. K. (2015b). "Tools in the Trunk," in *Doing Bayesian Data Analysis*. Editor J. K. Kruschke. Second EditionSecond edition edn (Boston: Academic Press), 721–736. doi:10.1016/B978-0-12-405888-0.00025-8

Kumar, R., Carroll, C., Hartikainen, A., and Martin, O. (2019). Arviz a Unified Library for Exploratory Analysis of Bayesian Models in python. *Joss* 4, 1143. doi:10.21105/joss.01143

Lelkes, Y., Krosnick, J. A., Marx, D. M., Judd, C. M., and Park, B. (2012). Complete Anonymity Compromises the Accuracy of Self-Reports. *J. Exp. Soc. Psychol.* 48, 1291–1299. doi:10.1016/j.jesp.2012.07.002

Linowes, J., and Schoen, M. (2016). *Cardboard VR Projects for Android*. Birmingham, United Kingdom: Packt Publishing Ltd.

Makowski, D., Ben-Shachar, M., and Lüdecke, D. (2019). bayestestR: Describing Effects and Their Uncertainty, Existence and Significance within the Bayesian Framework. *Joss* 4, 1541. doi:10.21105/joss.01541

Mallaro, S., Rahimian, P., O'Neal, E. E., Plumert, J. M., and Kearney, J. K. (2017). "A Comparison of Head-Mounted Displays vs. Large-Screen Displays for an Interactive Pedestrian Simulator," in Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology (New York, NY, USA: Association for Computing Machinery), 1–4. doi:10.1145/3139131.3139171*VRST*'17

Maruhn, P., Dietrich, A., Prasch, L., and Schneider, S. (2020). "Analyzing Pedestrian Behavior in Augmented Reality - Proof of Concept," in 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), 313–321. doi:10.1109/VR46266.2020.00051

Maruhn, P., and Hurst, S. (2022). "Effects of Avatars on Street Crossing Tasks in Virtual Reality," in Proceedings of the 21st Congress of the International Ergonomics Association (IEA 2021). Editors N. L. Black, P. W. Neumann, and I. Noy (Cham: Springer International Publishing). doi:10.1007/978-3-030-74614-8_26

McComas, J., MacKay, M., and Pivik, J. (2002). Effectiveness of Virtual Reality for Teaching Pedestrian Safety. *CyberPsychology Behav.* 5, 185–190. doi:10.1089/109493102760147150

McElreath, R. (2020). *Statistical Rethinking : A Bayesian Course with Examples in R and Stan*. Boca Raton: Chapman & Hall/CRC.

Meade, A. W., and Craig, S. B. (2012). Identifying Careless Responses in Survey Data. *Psychol. Methods* 17, 437–455. doi:10.1037/a0028085

Oswald, D., Sherratt, F., and Smith, S. (2014). Handling the hawthorne Effect: The Challenges Surrounding a Participant Observer. *Rev. Soc. Stud.* 1, 53–74. doi:10.21586/ross0000004

Peli, E. (1999). "Optometric and Perceptual Issues with Head-Mounted Displays," in *Visual Instrumentation: Optical Design and Engineering Principles*. Editor P. Mouroulis (New York: McGraw-Hill), 205–276. Available at: https://pelilab.partners.org/papers/Peli_Chapter%206_OptometricPercept_1999.pdf.

Prattico, F. G., Lamberti, F., Cannavo, A., Morra, L., and Montuschi, P. (2021). Comparing State-Of-The-Art and Emerging Augmented Reality Interfaces for Autonomous Vehicle-To-Pedestrian Communication. *IEEE Trans. Veh. Technol.*, 70, 1157–1168. doi:10.1109/TVT.2021.3054312

Renner, R. S., Velichkovsky, B. M., and Helmert, J. R. (2013). The Perception of Egocentric Distances in Virtual Environments - a Review. *ACM Comput. Surv.* 46, 1–40. doi:10.1145/2543581.2543590

Salvatier, J., Wiecki, T. V., and Fonnesbeck, C. (2016). Probabilistic Programming in python Using Pymc3. *PeerJ Comp. Sci.* 2, e55. doi:10.7717/peerj-cs.55

Schneider, S., and Bengler, K. (2020). Virtually the Same? Analysing Pedestrian Behaviour by Means of Virtual Reality. *Transportation Res. Part F: Traffic Psychol. Behav.* 68, 231–256. doi:10.1016/j.trf.2019.11.005

Schneider, S., Maruhn, P., Dang, N.-T., Pala, P., Cavallo, V., and Bengler, K. (2021). Pedestrian Crossing Decisions in Virtual Environments: Behavioral Validity in Caves and Head-Mounted Displays. *Hum. Factors* 0, 0018720820987446. doi:10.1177/0018720820987446

Schwebel, D. C., Gaines, J., and Severson, J. (2008). Validation of Virtual Reality as a Tool to Understand and Prevent Child Pedestrian Injury. *Accid. Anal. Prev.* 40, 1394–1400. doi:10.1016/j.aap.2008.03.005

Schwebel, D. C., Severson, J., He, Y., and McClure, L. A. (2017b). Virtual Reality by mobile Smartphone: Improving Child Pedestrian Safety. *Inj. Prev.* 23, 357. doi:10.1136/injuryprev-2016-042168

Schwebel, D. C., Severson, J., and He, Y. (2017a). Using Smartphone Technology to Deliver a Virtual Pedestrian Environment: Usability and Validation. *Virtual Reality* 21, 145–152. doi:10.1007/s10055-016-0304-x

Simpson, G., Johnston, L., and Richardson, M. (2003). An Investigation of Road Crossing in a Virtual Environment. *Accid. Anal. Prev.* 35, 787–796. doi:10.1016/S0001-4575(02)00081-7

Sobhani, A., Farooq, B., and Zhong, Z. (2017). “Distracted Pedestrians Crossing Behaviour: Application of Immersive Head Mounted Virtual Reality,” in 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), 1–6. doi:10.1109/ITSC.2017.8317769

Vehtari, A., Gelman, A., and Gabry, J. (2017). Practical Bayesian Model Evaluation Using Leave-One-Out Cross-Validation and Waic. *Stat. Comput.* 27, 1413–1432. doi:10.1007/s11222-016-9696-4

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., and Bürkner, P.-C. (2021). Rank-Normalization, Folding, and Localization: An Improved $\hat{R}$ for Assessing Convergence of MCMC (With Discussion). *Bayesian Anal.* 16, 1. doi:10.1214/20-ba1221

Woldegiorgis, B. H., Lin, C. J., and Liang, W.-Z. (2018). Impact of Parallax and Interpupillary Distance on Size Judgment Performances of Virtual Objects in Stereoscopic Displays. *Ergonomics* 62, 76–87. doi:10.1080/00140139.2018.1526328

Wynne, R. A., Beanland, V., and Salmon, P. M. (2019). Systematic Review of Driving Simulator Validation Studies. *Saf. Sci.* 117, 138–151. doi:10.1016/j.ssci.2019.04.004

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

TUM