Frontiers in Virtual Reality

ORIGINAL RESEARCH
published: 01 July 2022
doi: 10.3389/frvir.2022.697934

The Potential of Immersive Virtual Reality for the Study of Event Perception

Julia Misersky[1]*, David Peeters[2] and Monique Flecken[1,3,4]

[1]Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands, [2]Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands and Department of Communication and Cognition, TiCC, Tilburg University, Tilburg, Netherlands, [3]Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands and Amsterdam Centre for Language and Communication, University of Amsterdam, Nijmegen, Netherlands, [4]Amsterdam Centre for Language and Communication, University of Amsterdam, Amsterdam, Netherlands

In everyday life, we actively engage in different activities from a first-person perspective. However, experimental psychological research in the field of event perception is often limited to relatively passive, third-person computer-based paradigms. In the present study, we tested the feasibility of using immersive virtual reality in combination with eye tracking with participants in active motion. Behavioral research has shown that speakers of aspectual and non-aspectual languages attend to goals (endpoints) in motion events differently, with speakers of non-aspectual languages showing relatively more attention to goals (endpoint bias). In the current study, native speakers of German (non-aspectual) and English (aspectual) walked on a treadmill across 3-D terrains in VR, while their eye gaze was continuously tracked. Participants encountered landmark objects on the side of the road, and potential endpoint objects at the end of it. Using growth curve analysis to analyze fixation patterns over time, we found no differences in eye gaze behavior between German and English speakers. This absence of cross-linguistic differences was also observed in behavioral tasks with the same participants. Methodologically, based on the quality of the data, we conclude that our dynamic eye-tracking setup can be reliably used to study what people look at while moving through rich and dynamic environments that resemble the real world.

Keywords: virtual reality, event perception, eye tracking, cave, motion events

OPEN ACCESS

Edited by:
*David Swapp,
University College London,
United Kingdom*

Reviewed by:
*Harry Farmer,
University of Greenwich,
United Kingdom
Thomas Schubert,
University of Oslo, Norway*

****Correspondence:***
*Julia Misersky
julia.misersky@mpi.nl*

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

Received: *20 April 2021*
Accepted: *24 May 2022*
Published: *01 July 2022*

Citation:
*Misersky J, Peeters D and Flecken M
(2022) The Potential of Immersive
Virtual Reality for the Study of
Event Perception.
Front. Virtual Real. 3:697934.
doi: 10.3389/frvir.2022.697934*

1 INTRODUCTION

Over the past decades, experimental research into human language and cognition has greatly benefited from the use of computer monitors to display carefully developed experimental stimuli to participants. The use of strict experimental designs implemented through computer paradigms has allowed for the advancement of cognitive and psycholinguistic theory in unprecedented ways. Recently, however, in several subfields of the experimental study of language, perception, and cognition, more and more attention is devoted towards attempting to develop paradigms and methods that explicitly combine high experimental control with high ecological validity (e.g., Hari et al., 2015; Knoeferle, 2015; Willems, 2015; Blanco-Elorrieta and Pylkkänen, 2018; Peeters, 2019). Crucially, unlike in the typical traditional experimental setup that places individual participants in front of a computer monitor, in real-life settings "people are participating in the events of their world, and they do not only serve as passive observers" (Hari et al., 2015, p. 184). In this view, by employing passive and highly controlled experimental stimuli, some traditional studies and their interpretations may potentially have an intrinsically limited theoretical scope, as human behavior in everyday life often takes place in rather dynamic and multidimensional settings. This is particularly relevant to the study of event cognition, which is concerned with how we perceive and understand everyday events such as walking to the bus or making a coffee. It is unclear whether theories and models that were mainly built on the basis of traditional, strictly controlled computer paradigms fully generalize to people's everyday behavior in the real world.

Immersive virtual reality is a rapidly advancing technology that offers researchers the opportunity to develop experimental paradigms that combine high ecological validity with high experimental control, thereby allowing for the experimental study of human cognition and behaviour under rich and dynamic circumstances (Blascovich et al., 2002; Bohil et al., 2011; Parsons, 2015; Pan and Hamilton, 2018; Peeters, 2019). In the current study, we investigate the potential and feasibility of using immersive virtual reality (VR) in a Cave Automatic Virtual Environment (CAVE) to reliably collect looking behavior in participants that move through dynamic virtual environments. In a CAVE setup, participants are surrounded by large screens or walls on which 3-D environments are projected. These virtual surroundings typically adapt to the continuously tracked behavior (e.g., head movement, eye gaze) of the actively engaged participant (Cruz-Neira et al., 1993). As such, the CAVE setup can be used to create a relatively realistic setting that may mimic the dynamics and multimodal richness of day-to-day situations. This hence potentially allows the experimental researcher to step away from computer-based tasks while still exerting the required control over the presented stimuli and the collection of various types of theoretically informative data.

Recently, a number of different studies have successfully used immersive VR as a method to study fundamental aspects of human language and cognition (for overview, see Peeters, 2019). For instance, VR and eye tracking have been combined to gain clearer insights into how listeners may predict upcoming words for visual objects in visually rich settings (Heyselaar et al., 2020). In general, virtual reality has been successfully used to study a wide variety of language-related topics, including syntactic priming, audiovisual integration, language evolution, bilingualism, reading, and the role of eye gaze in social interaction (e.g., Heyselaar et al., 2015; Hömke et al., 2018; Peeters and Dijkstra, 2018; Tromp et al., 2018; Cañigueral and Hamilton, 2019; Mirault et al., 2020; Nölle et al., 2020). It is thus not unlikely that it will soon become one of the by default available options as a method of stimulus display in the experimental researcher's toolbox in the study of language and cognition. However, while in these earlier studies the virtual environments typically had a substantial degree of dynamicity and interactivity, participants themselves remained relatively static throughout the experiment. In contrast, many aspects of our lives require our dynamic participation, especially when it comes to the day-to-day events we experience. The current study therefore aims to move away from 3-D settings in which participants remain static, and test the feasibility of putting participants into motion within a dynamic VR environment, while collecting informative eye-tracking data to inform theories of language and cognition.

As a relevant testing ground, we opted to focus on the domain of event cognition. Previous psycholinguistic research employing computer-based paradigms in this field have studied how people perceive and make sense of events, for instance in relation to the structural properties that their native language has on offer. The timing of an event (i.e., its beginning and end) may, for example, be made explicit linguistically through grammatical aspect marking. Specifically, progressive aspect (e.g. in English, She was dancing) can be used to highlight the inner temporal structure and continuousness of an event (Klein, 1994). In contrast, perfective aspect (e.g., in English, She danced) focuses on the event in its entirety, without specific details about the inner structure of the event itself. To study the relation between language and cognition in the case of event perception, typical experimental studies in this domain show participants videos on a computer monitor while collecting eye-tracking data, in the presence or absence of them verbally producing event descriptions, for instance testing whether different types of aspect marking influence event perception. Theoretical accounts thus acknowledge that language in general, and aspect in particular, can guide how we make sense of events, regardless of whether we observe others as agents in events or whether we are the agents ourselves (e.g., Magliano et al., 2014; Swallow et al., 2018). However, in everyday life, we tend to experience events in 3D and from an active, first-person perspective, which contrasts with the (2D, relatively passive) experimental paradigms commonly used in this research field.

Importantly, languages differ in the ways in which aspect is expressed, and hence people may indeed attend to different features within a scene depending on their language background. This hypothesis has been studied cross-linguistically, for instance by comparing speakers of aspectual languages (e.g., English, Modern Standard Arabic) to speakers of non-aspectual languages (e.g., German, Swedish; Von Stutterheim et al., 2012; Athanasopoulos and Bylund, 2013; Flecken et al., 2014). Watching video clips of motion events, participants have been asked to describe the events in one sentence, while their eye movements were tracked throughout (Von Stutterheim et al., 2012; Flecken et al., 2014). Cross-linguistic differences were found for video clips in which agents approached but did not reach a potential goal (e.g., a person running towards the train station, but the video ending before the person running actually reached it). Specifically, the data revealed that speakers of non-aspectual languages (e.g., German) were more likely to mention a goal (i.e., the train station) in their descriptions. In addition, speakers of non-aspectual languages looked more often and longer towards the goal, compared to speakers of aspectual languages (e.g., English, Modern Standard Arabic). The cross-linguistic difference in verbal motion event descriptions has further been replicated in a comparison of English (aspectual) versus Swedish (non-aspectual) speakers (Athanasopoulos and Bylund, 2013) and confirmed by an event-similarity-judgment task performed by a new set of participants (see *Methods* section below for details).

In sum, the above results suggest that speakers of non-aspectual languages (e.g., German) are more likely to show a bias towards potential goals or endpoints when watching others in motion events, whereas speakers of aspectual languages (e.g., English) do not show such a bias. This cross-linguistic difference is particularly apparent in verbal tasks, and seems to be less pronounced in non-verbal tasks (Athanasopoulos and Bylund, 2013, who tested event similarity judgment in verbal and nonverbal experimental setups). Following up on this work, we will place native speakers of German (a non-aspectual language) and English (an aspectual language) on a treadmill in an immersive 3D environment to study looking patterns in a rich and dynamic setting. As such, we set out to collect data that was at the same time methodologically and theoretically informative, testing whether this novel method allows reliable data collection and, if so, whether earlier theories generalize to richer, dynamic environments.

The main methodological aim of the current study was thus to test the feasibility of combining immersive VR with eye tracking in moving participants. If reliable data can be collected in such a setup, this will open up a wide range of possibilities for future studies investigating the relation between language and perception, for instance in the domain of (motion) event cognition and visual perception. In line with previous research (e.g., Von Stutterheim et al., 2012; Flecken et al., 2014; Flecken et al., 2015), we employed a design in which participants could observe landmark objects as well as potential endpoints in a scene, and were interested in their looking behavior to the latter as a function of their native language background. In general, if reliable data collection is feasible in this experimental setup, we would expect a significant increase in looks to endpoints over time in both groups as they approached the endpoint. After all, an endpoint that is approached and thus comes to take up an increasingly larger proportion of the overall visual scene would naturally attract relatively more looks over time across both participant groups. More specifically, over and above this hypothesized main effect, we expected Germans to fixate on endpoints more compared to English speakers in general. Unlike previous studies, participants were not observing the scenes passively, but actively walked through the scenes while encountering objects and endpoints along the way. To keep the setup as similar to real-life motion events as possible, participants were not prompted to verbalize what they were experiencing during the VR task. In the current study, participants did not observe others in the role of an agent, but became agents themselves. To be able to compare this novel setup to previous work using the third-person perspective, we also utilized verbal behavioural tasks on event cognition, which have resulted in cross-linguistic differences in computer-based paradigms in earlier work (Von Stutterheim et al., 2012; Athanasopoulos and Bylund, 2013; Flecken et al., 2014). Note that previous behavioural work has shown that when viewing continuous events from a first-person perspective, participants build event models based on similar attributes as when they build event models when observing others (Magliano et al., 2014; Swallow et al., 2018).

2 MATERIALS AND METHODS

2.1 Participants

Twenty-four speakers of German (15 female, 20–31 years of age, M = 22.92, SD = 2.84), and 24 native speakers of English (14 female, 18–35 years of age, M = 23.92, SD = 4.6) participated in this experiment.[1]

Participants were invited *via* the online participant database of the Max Planck Institute for Psycholinguistics and advertising through flyers and social media. They gave written informed consent prior to their participation and received a monetary compensation for their time (10 Euros per hour). The experiment was approved by the ethics committee of the Social Sciences Faculty of Radboud University, Nijmegen, Netherlands.

After completing the experiments, we asked participants to provide information on their educational and language background. Specifically, we asked about other countries they have lived in (for a minimum of 3 months) and which other language apart from their native they were most proficient in. In addition, we asked them to rate their levels of speaking, writing, comprehension, and reading in their most proficient L2 (poor, sufficient, good, very good, or excellent). Overall, English native speakers came from a greater variety of home countries (incl. Australia, Canada, Luxemburg, Indonesia, United Kingdom, United States, and Trinidad) compared to German speakers (Germany). English speakers frequently reported poor or sufficient skills in their most proficient L2 (e.g., French, Spanish, German) whereas German speakers most frequently named English and Dutch as their most proficient L2 with skills predominantly rated as very good and excellent. Note also that the majority of all participants were international students at Radboud University, where study programs were held in English, meaning that the German speakers were often exposed to a language other than their L1, and frequently used this/these language(s) in daily life, too.

2.2 VR Material Selection and Trial Setup

In the VR part of the experiment, participants walked through a total of 48 trials on a treadmill. Four 3-D road types were designed: a parkland lane, an urban road, a sandy countryside path, and a forest trail (see **Figure 1**). Fifty-two unique 3-D objects were selected for a pre-test. Using a paper-and-pencil questionnaire, 18 native German speakers (9 female, 25–41 years of age, M = 30.11, SD = 4.68) and 20 English speakers (10 female, 21–37 years of age, M = 27.8, SD = 3.64) first named a 2-D picture

[1] In addition, one English speaker participated but was excluded, as they were raised in both English and German. Further, nine German-speaking participants took part but were excluded from further analysis due to technical issues resulting in the incorrect presentation of stimulus lists. The final VR analysis was thus based on 24 participants in each language group (for further exclusion in the behavioural tasks see the Results section).

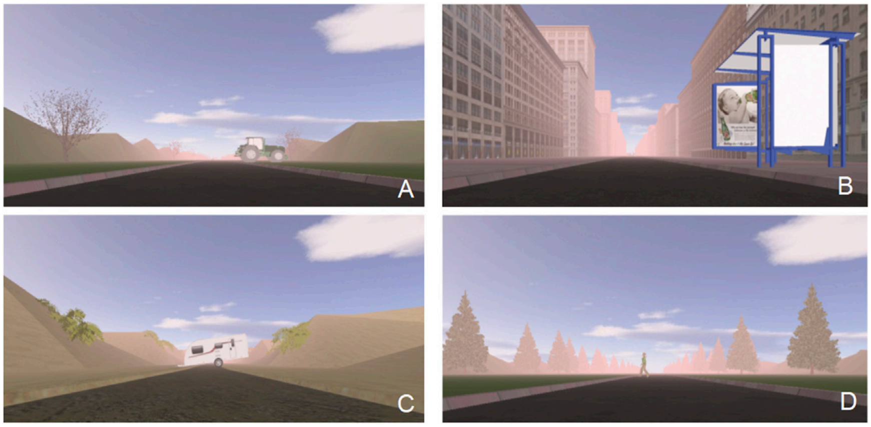

FIGURE 1 | Left to right: Examples of the four road types and objects within them; **(A)** parkland lane with LM (a tractor) on the right, **(B)** urban road with LM (bus stop) on the right; **(C)** sandy countryside path with EP (camper trailer) at the end of the road, **(D)** forest trial with a virtual agent crossing the road.

of each object, and then rated their prototypicality on a 7-point scale (1 = not prototypical at all, 7 = very prototypical). None of these participants took part in the main experiment. For the experimental trials of the main experiment, the 48 objects with highest name agreements and prototypicality ratings (mean prototypicality ratings ranging from 5.1 to 6.8 for the German speakers, and from 4.9 to 6.8 for the English speakers) within each language were chosen. Forty of the objects were specifically designed for this experiment by a graphics designer using Autodesk Maya software. The remaining objects (n = 8) were taken from the standardized database of 3-D objects provided by Peeters (2018).

Twenty-four of the trials in the main experiment were experimental trials, in which participants always encountered two 3-D objects, one on the side of the road (landmark, LM) and one at the end of the road (endpoint, EP). Unbeknownst and invisible to participants, each trial was split into three phases. In Phase 1, only the LM was visible. As the trial continued, participants entered Phase 2, in which both LM and EP were in view. As they passed the LM, participants entered Phase 3, in which only the EP was visible. Half of the experimental trials (n = 12) stopped before the EP was reached (short trial), whereas in the other half, the trial stopped as they arrived at the EP (long trial). This ensured that participants could not anticipate whether or not they would actually reach the goal (EP object), similar to previous behavioral studies based on video clips (Von Stutterheim et al., 2012; Flecken et al., 2014). Importantly, previous cross-linguistic differences were obtained only for events in which actors did not reach a goal; this was the critical condition leaving room for variability in specifying and looking at potential goals. Here, Phase 2 would correspond to a situation in which a potential goal is visible, but whether or not it would be reached was unclear yet (regardless of short/long trials).

The objects were matched to appear always with the same road type (see **Supplementary Appendix S1** for a list of all objects and the corresponding road type). For each road type, the items were balanced across participants such that an object presentation was balanced across long and short trials, in their position in LM or EP location, and whether they appeared on the left or right side of the road in the case of LM. This resulted in eight lists across participants, in which trials were pseudo-randomized for each participant. In the filler trials (n = 24), participants encountered a virtual agent crossing the road in front of them. Four virtual agents (two female, two male) were adapted from existing stock avatars produced by WorldViz (Vizard, Floating Client 5.4, WorldViz LLC, Santa Barbara, CA). See **Figure 1** for an example of the trial setup and stimuli.

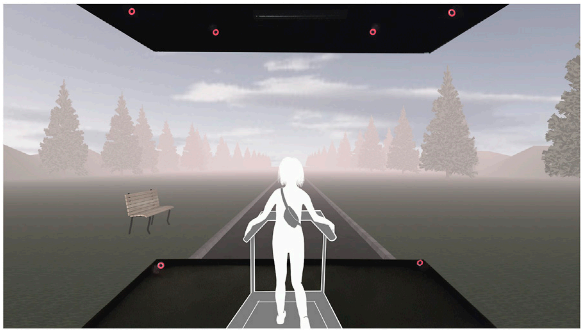

FIGURE 2 | Illustration of the CAVE setup with its three screens. Participants walked on a treadmill in a variety of dynamic virtual environments, encountering both landmark (here: a bench) and endpoint objects along their way while their eye movements were continuously tracked and recorded. A subset of the infrared-cameras is marked red.

2.3 Apparatus

2.3.1 The CAVE System

The CAVE system is made up of three screens (255 × 330 cm, VISCON GmbH, NeukirchenVluyn, Germany) that were arranged to the left, to the right and in front of the participant, as illustrated in the in **Figure 2**. Each of the screens was illuminated through a mirror by two projectors (F50, Barco N.V., Kortrijk, Belgium). The two projectors showed two vertically displaced images, which overlapped in the middle of each screen, meaning the display was only visible as the combined overlay of the two projections. Infrared motion capture cameras (Bonita 10, Vicon Motion Systems Ltd., United Kingdom) and the Tracker 3 software (Vicon Motion Systems Ltd., United Kingdom) allowed for optical tracking. The infrared cameras detected the positions of retroreflective markers mounted onto the 3-D glasses by optical–passive motion capture (see below for further details). A total of ten infrared cameras were placed around the edges of the screens in the CAVE: Six cameras were positioned at the upper edges, and four cameras at the bottom edges. All cameras were oriented toward the middle of the CAVE system, where the participants were located during testing. The positions of a subset of the cameras are indicated in **Figure 2**. The resolution of the CAVE system was 2560 × 1956 pixels per screen and the refresh rate was 60 Hz, which allowed for the glasses and CAVE system to be in sync. This CAVE setup is described in further detail in Eichert et al. (2018). Unlike earlier studies using this CAVE lab, a treadmill was placed in the centre of the system, such that during walking, all three screens covered participants' entire horizontal visual field. The eyes of the participant were approximately 180 cm away from the middle screen.

From the control room, the experimenter could see the participant and the displays on the screens through a large window behind the participant (thus facing the central CAVE screen, similar to the view depicted in **Figure 2**).

The experiment was programmed in Python, and run using 3-D application software (Vizard, Floating Client 5.4, WorldViz LLC, Santa Barbara, CA). To allow for a realistic experience, we made sure that moving forward a "virtual" metre inside the VR environment corresponded to moving forward a "real" metre on the treadmill. Note that object sizes were dynamic, i.e., perceived object size changed throughout a trial as the participant advanced on the virtual terrain—as in real life objects became relatively larger when participants approached them. To further enhance the naturalness of the walking experience, bird and wind sounds were presented through speakers located in the CAVE (Logitech Surround Sound speaker system Z 906 5.1).

2.3.2 Eye Tracking

Eye tracking was performed using glasses (SMI Eye-Tracking Glasses 2 Wireless, SensoMotoric Instruments GmbH, Teltow, Germany) that combine the recording of eye gaze with the 3-D presentation of VR through shutter glasses. The recording interface was based on a tablet (Samsung Galaxy Note 4), which was connected to the glasses by cable. The recorder communicates with the externally controlled tracking system *via* a wireless local area network (WIFI), allowing for live data streaming. The glasses were equipped with a camera for binocular 60-Hz recordings and automatic parallax compensation. The shutter device and the Samsung Galaxy 4 tablet were placed in a small shoulder bag that participants carried on their back during walking on the treadmill (see **Figure 2**). Gaze tracking accuracy was estimated by the manufacturer to be 0.5° over all distances. The delay of the eye-tracking signal (i.e., the time it takes for the eye-tracking coordinates to reach the VR computer and its output file) was estimated to be 60 ± 20 ms. In addition, we combined the eye tracking with optical head tracking. Optical head tracking was accomplished by placing spherical reflectors on the glasses. We were thus able to identify the exact location of the eye gaze in three spatial dimensions (X, Y, and Z). This in turn allowed for an immersive experience in the VR, as the presentation of the 3-D world moved in accordance with the participants' (head) motion on the treadmill. In order to achieve smooth 3-D presentation, the 3-D eye-tracking glasses were equipped with reflectors (three linked spheres) magnetically attached to both sides of the glasses, which worked as passive markers that were detected by the infrared tracking system in the CAVE. The tracking system was trained to the specific geometric structure of the markers and detected the position of the glasses with an accuracy of 0.5 mm.

Calibration of the eye tracker was carried out in two steps. In an initial step, general tracking of the pupils was tested in the control room using the Samsung Galaxy 4 tablet. If this was successful, shutter device and tablet were stowed into the shoulder bag, and a second calibration step was carried out in the 3-D environment as the participant was walking on the treadmill. For this step, we used a virtual test scenery, resembling a tea house in which three differently colored spheres were displayed in front of the participants. The position of the three spheres differed in all three spatial coordinates. Participants were asked to look at the three displayed spheres successively, which the experimenter communicated to the participants *via* the microphone. The computer software computed a single dimensionless error measure of the eye tracker combining the deviation in all three coordinates for all three spheres. This second calibration step was repeated until a minimal error value (<5° difference between the invisible vector between the shutter glasses and the centre of each sphere and the invisible vector between the shutter glasses and where a participant's fixation was actually estimated to be by the system), and thus maximal accuracy, was reached.

2.3.3 Regions of Interest

To determine target fixations, we defined individual 3-D regions of interest (ROIs) around each object in the virtual space. ROIs are defined by a rectangular prism that encloses the object. The X (width) and Y (height) dimensions of the ROI were adopted from the frontal plane of the object's individual bounding box, facing the participant. In the experiment software, eye gaze towards an object was detected if the line of sight collided with an object's ROI in the scene. In other words, the eye-tracking software automatically detected when the eye gaze was directed at one of the ROIs and coded the information online in the data stream. Note that the dimensions of width, height, and depth of a given ROI do not change, but the position of the participant in the virtual world changes because of the forward (Z-axis) motion. This can be thought of as an analogy to motion in the real world: The absolute values of height, width and depth of, for example, a parked car do not change; the car is always the same size. However, as you are approaching the parked car, your line of sight adjusts so you can see the top, front or side of the car.

2.3.4 Data Processing

Based on the inspection of the output data files, our data were cleaned before further processing. Each row in the output signified one frame of ~16.6 m s. In some cases, the data showed timestamp duplicates. If a timestamp occurred more than once within a participant and trial, all frames with the duplicate were marked. Further, freezing during eye tracking also incidentally occurred, meaning that the coordinates for the X, Y, and Z planes showed duplicates as well. When simultaneous freezing across all three coordinates was observed more than once within a participant and trial, all frames with the duplicate were marked. Marked frames were considered unreliable data points. Overall, 14.96% of the data were affected by duplicates in Phase 1, 11.22% in Phase 2, and 26.07% in Phase 3. The data were binned prior to analysis, with each bin containing three frames. Any bin containing at least one marked frame was excluded from the final analysis. A fixation was then defined as saccades towards an ROI for at least six frames or ~100 ms. Shorter saccades were considered as unlikely to represent a fixation (cf. Eichert et al., 2018).

2.3.5 VR Eye-Tracking Data Analysis

The eye-tracking data obtained in the VR experiment was analyzed using growth curve analysis (GCA) on the cleaned data (see above). GCA uses a linear mixed regression approach and has been successfully used for visual world paradigms (VWPs) before (see Mirman, 2016). GCA can give insights into whether looking behavior differs between groups or items within a given time window, and thus allowed us to see whether our Language groups differed with regards to their fixations on EP in the VR scenes. Thus, we focused our analysis on the phases when the EP is visible, namely Phases 2 and 3. Unlike conventional VWPs, our study used a free viewing setup. As such, there were no fixed time points from which we were able to restrict our time windows of interest. Small changes–depending on the setup—can impact the results and their interpretation (Peelle and Van Engen, 2020). That said, there are no specific recommendations for the choice of time-window length, and transparency regarding the chosen approach is hence key. To keep the bias in choosing the time-windows to a minimum, we opted to use the same time-window length for both Phases 2 and 3. Combining visual inspection (i.e., when fixations start diverging in Phase 2) and information about Phase length in our setup (i.e., the maximal length of Phase 3 in short trials being 5 s), we opted to restrict our analysis to the last 5 s of Phase 2, and the first 5 s of Phase 3.[2]

TABLE 1 | Output of the model on EP fixations in Phase 2 (glmer (Hits ~ Language * (Time^1 + Time^2 + Time^3) + Nuisance + (1 | Item-pair) + (1 | Participant)).

Fixed effects	β Estimate	SE	z-value
Intercept	−4.057	0.185	−21.933***
LanguageEng	−0.048	0.229	−0.210
Time^1	3.551	0.216	16.441***
Time^2	−0.048	0.215	−0.224
Time^3	−0.762	0.210	−3.624***
Nuisance	1.837	0.009	210.835***
LanguageEng: Time^1	−0.049	0.294	−0.168
LanguageEng: Time^2	−0.463	0.292	−1.586
LanguageEng: Time^3	−0.051	0.287	−0.178

***p < 0.001; **p < 0.01; *p < 0.05.

TABLE 2 | Output of the model on EP fixations in Phase 3 (glmer (Hits ~ Language * (Time^1 + Time^2 + Time^3) + Nuisance + (1 + Time^1 + Time^2 | Item-pair) + (1 + Time^1 + Time^2 + Time^3 | Participant)).

Fixed effects	β Estimate	SE	z-value
Intercept	−1.826	0.141	−12.929***
LanguageEng	−0.161	0.162	−0.996
Time^1	1.438	0.258	5.576***
Time^2	−0.543	0.253	−2.148*
Time^3	0.187	0.218	0.858
Nuisance	2.585	0.009	274.799***
LanguageEng: Time^1	−0.294	0.320	−0.918
LanguageEng: Time^2	0.678	0.325	2.084*
LanguageEng: Time^3	−0.139	0.307	−0.453

***p < 0.001; **p < 0.01; *p < 0.05.

GCA uses polynomial orthogonal time terms as predictors in the linear mixed regression model, which describe the shape of the fixation curves in our time windows of choice. For both Phase 2 and 3 we chose third-order polynomial time terms as predictors in the respective models. Each time term describes a different aspect of the fixation curve (cf. Sauppe, 2017): Time^1 (linear) describes the angle of the fixation curves. Time^2 (quadratic) describes the rate of increase or decrease. Time^3 describes earlier or later increases or decreases of the fixation curves. We ran generalized linear mixed effect models on the Endpoint hits by Language (treatment-coded) interacting with the Time predictors as fixed effects. In addition, we introduced a fixed effect of a Nuisance predictor (Sassenhagen and Alday, 2016), which took into account all Endpoint hits in the previous bin to use as a predictor for the current bin in order to reduce temporal autocorrelation in the continuous eye-tracking data. The maximal random effects structure allowing for convergence was used for all models (see full models in **Tables 1**, **2**). All models were computed using the lme4 package (Version1.1-21; Bates et al., 2015) in R (Version 3.6.0; R Core Team, 2019).

In terms of our methodological focus, we would expect a linear increase in looks towards objects, and EPs in particular, as trial time goes on. This is based on the setup of the study, in which the participants move along the road with objects appearing to get closer. In Phases 2 and 3 especially, the EP thus increasingly occupies the view field of the participants such that looks to EPs should become more frequent as each trial continues. In line with

[2] The use of time-windows of 5 s was chosen for the following reasons: In line with our interest in views towards the EP, we wanted to analyse data for those time-windows in which participants had a free choice to look towards either the EP or something else. Recall that in Phase 2, participants had this choice throughout, whereas in Phase 3, participants had free choice of looking or not looking towards the EP only in the beginning of the Phase (coinciding with the end of short trials). We opted to take an objective and conservative approach to the data analysis, and thus wanted to use the same time-window length for both Phases 2 and 3. Please consult **Supplementary Appendix S2** for an additional analysis of Phase 2 using a shorter time-window and yielding comparable results.

our research question regarding cross-linguistic differences, interactions of any of the time terms with our Language predictor would indicate differences in EP fixation behavior between the German and English speakers.

2.4 Procedure

After receiving written information about the experiment and giving consent, every experimental session started with the VR experiment. Participants put on the VR glasses, which were fastened with a drawstring strap. The calibration was then carried out, and when successful, all equipment powering the glasses was stored in a small shoulder bag, which the participant was instructed to wear throughout the experiment. The participant was then led into the room with the CAVE system and asked to stand on the treadmill. The second calibration step was then carried out, ensuring accurate tracking of the eye gaze and simultaneously checking whether the participant could see the projections in 3-D. The treadmill was turned on and set to a fixed, comfortable walking speed (~3 km per hour) to match the presentation of the moving environment. The experimenter then explained the task of the experiment, which was to walk on the treadmill and listen out for bird sounds. Upon hearing a bird, the participant had to press a button on the handle of the treadmill. The participant experienced four practice trials (one of each road type, LM on the left half of the trials, equal amounts of long and short trials) to get used to this setup, and had the opportunity to ask questions afterwards. The experiment proper then started with another calibration in the CAVE. After half the trials, the participant had a self-paced break but had to remain on the treadmill. The experimenter re-checked the calibration, and re-calibrated where necessary, then started the second half of the experiment when the participant was ready. After completion of the VR experiment, the experimenter helped the participant to step off the treadmill and remove the VR glasses.

Three computer-based tasks followed, all of which were programmed using Presentation software (Neurobehavioral Systems).

First, participants carried out an object recognition memory task as a test of whether they paid attention during the VR experiment. On a white computer screen, participants saw a fixation cross, then an object for 1200 ms, which was followed by the question whether they had seen it previously in the VR experiment. Participants had to indicate their decision *via* a button press on a button box. Overall, each participant saw 24 previously seen objects (12 as EPs, 12 as LMs), and 24 new objects. The order of objects was randomized. High accuracy overall in this task was hypothesized to indicate participants had paid attention to the objects in the task.

Second, participants performed an event description task (Von Stutterheim et al., 2012; Flecken et al., 2014), in which on each trial they saw a fixation cross followed by a video clip. They were asked to describe what was happening in the clip using a single sentence. Participants were instructed to speak into a small microphone, and were allowed to speak as soon as they felt ready to describe what is happening. They were asked not to focus on details (e.g., colors, backgrounds, such as “the sky is blue”) and just on the event itself. A blank followed, and the fixation cross re-appeared, upon which participants were able to start the next clip. Overall, each participant saw 50 video clips in a randomized order. In line with the previous work, our clips of interest were those in which a potential goal was not reached. For those clips ($n = 10$), German speakers were expected to mention EPs more often compared to English speakers.

Third, participants were instructed to carry out a similarity judgment task (adapted from Athanasopoulos and Bylund, 2013, their Experiment 2a). In each trial, they saw a triad of three consecutive short video clips, labelled A, B, and X. X always showed intermediate goal orientation, meaning there was a possible endpoint but no arrival was shown. The alternates (A and B) showed low (i.e., motion along a trajectory, no immediate endpoint) and high goal orientation (i.e., arrival at endpoint was shown) respectively. Their task was to judge whether clip X was more similar to clip A or to clip B. To indicate their decision, they had to press one of two buttons. In total, they saw 38 triads in a randomized order. German participants were expected to show higher endpoint bias (i.e., rating X as more similar to the high goal alternate) compared to English speakers.

Lastly, participants filled in a paper questionnaire regarding their language and educational background (see above for details). The experimenter thanked them for their time and debriefed them regarding the purpose of the study.

All interactions between the experimenter and the participants were carried out in the participants’ native language. An experimental session took between 90 and 120 min.

3 RESULTS

3.1 VR Experiment

3.1.1 Phase 2: Both LM and EP are Visible

Figure 3 shows participants’ fixations to the EP in all of Phase 2. For the GCA analysis, we focused on the last 5 s of Phase 2, the results of which can be found in **Table 1**.

The significant main effect for Time^1 ($p < 0.001$), here with a positive β Estimate, describes an overall increase in looks towards the EP over time, which is to be expected as participants approached the EP in this phase. The negative β Estimate of the main effect for Time^3 ($p < 0.001$) describes the “S-shape” of the fixation curve, which is not pronounced. No interactions between the orthogonal Time terms and Language were observed, suggesting no statistical differences in looking behavior (EP fixations) across the two Language groups in this Phase.

3.1.2 Phase 3: Only EP is Visible

Figure 4 shows participants’ fixations to the EP in the first 5 s of Phase 3, while **Table 2** shows the results for the GCA analysis for this time-window. Like in Phase 2, we observed a significant main effect of Time^1 ($p < 0.001$) with a positive β Estimate, meaning overall looks towards the EP increased over time as to be expected. The negative β Estimate of marginally significant main effect of Time^2 ($p = 0.032$) reveals that the rate of increase was low. Most importantly, there was a marginal interaction between English and Time^2 ($p = 0.037$) with a positive β Estimate value, suggesting English speakers showed a slightly stronger rate of increase in fixations towards the Endpoint object in this Phase. This cross-linguistic difference was unexpected.

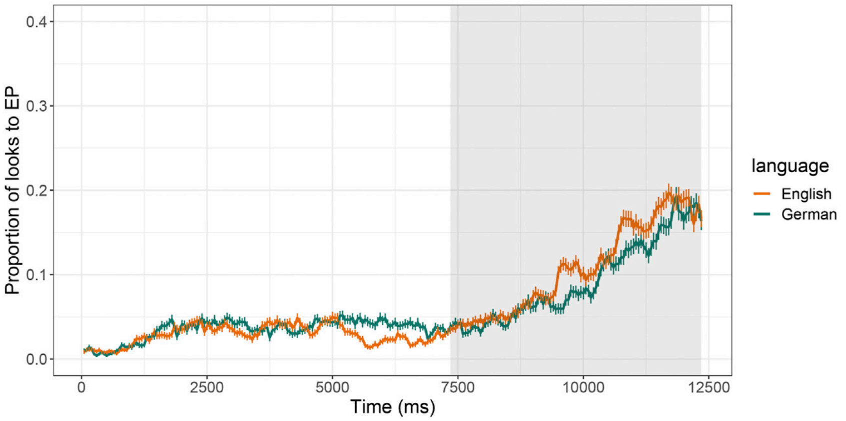

FIGURE 3 | Proportion of looks (fixations) to the EP in Phase 2 (both LM and EP visible) by Language, with the analysis time-window of 5 s highlighted in grey.

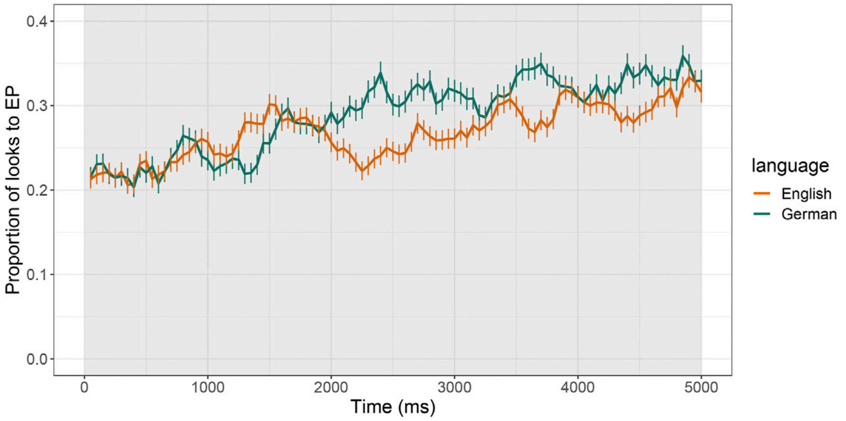

FIGURE 4 | Proportion of looks (fixations) to the EP in Phase 3 (only EP visible, collapsed for long and short trials) by Language, with the analysis time-window of 5 s highlighted in grey.

3.2 Object Memory Task

An analysis of the object memory task was performed for the 24 German speakers and 24 English speakers that successfully took part in the main VR experiment. Both English (M = 0.860, SD = 0.347) as well as German-speaking (M = 0.859, SD = 0.349) participants' accuracy was well above chance in this task, with no significant difference ($p = 0.9$) between the two groups, suggesting all participants were attentive to the objects in the VR scenes.

3.3 Event Description Task

Based on the data that was included in the VR analysis, technical errors or reported familiarity with the event description task, a small number of data sets were excluded. The analysis of the event description task was thus based on 22 German speakers and 24 English speakers, and focuses on the subset of the trials in which an EP was not reached ($n = 10$). A score of ten indicated an EP was mentioned in all clips, whereas a score of zero indicated no EP was mentioned for any of the clips. Means of EP mentions were similar across English (M = 4.167, SD = 1.993) and German speakers (M = 4.091, SD = 1.925). We ran a logistic regression model on the binomial data of score (0 = EP not mentioned within a trial, 1 = EP mentioned within a trial), with a fixed effect for Language, and random effects for Participant and Video Clip. Results indicate there were no differences in EP mentions between the Language groups (Intercept: β Estimate = -0.36505; z-value = -0.491; Language: β Estimate = 0.05648; SE = 0.38980, z-value = 0.145, p = 0.885).

3.4 Similarity Judgment Task

Based on the data included in the VR analysis and technical errors, a small number of data sets were excluded. The analysis of the event description task was thus based on 21 German speakers and 23 English speakers. Means for a bias towards matching clips based on their degree of endpoint-orientation were similar across English (M = 0.272, SD = 0.445) and German speakers (M = 0.247, SD = 0.431). We ran a logistic regression model on the binomial data of Bias (0 = no bias, 1 = strong bias) by Language, with random effects for Participant and Video Clip. Results indicate there were no differences in similarity judgments between the Language groups (Intercept: β Estimate = −1.1971; SE = 0.3694; Language: β Estimate = 0.1688; SE = 0.1417, z-value = −1.191, p = 0.23).

4 DISCUSSION

The current study set out to establish whether immersive virtual reality can be used reliably with moving participants in dynamic 3D settings to study aspects of human language and cognition. Two groups of participants (native speakers of English and native speakers of German) walked through a variety of virtual environments in a CAVE environment while their eye movements were continuously tracked. Both groups consistently looked at a wide variety of objects placed along the road and at a distance, and their proportion of looks at objects naturally increased while approaching them. Based on the observed significant overall increase in looks to the endpoints over time, we conclude that this setup can be reliably used to study what participants look at while immersed in rich and dynamic environments that resemble the real world.

In recent years, several studies have used virtual reality (VR) as a method to study various theoretically interesting characteristics of everyday human language and cognition in the lab (Pan and Hamilton, 2018; Peeters, 2019; Krohn et al., 2020). With regards to psycholinguistic research, VR has for instance been successfully implemented in studies on syntactic priming and subtle aspects of social interaction (e.g., Heyselaar et al., 2015; Hömke et al., 2018). In addition, previous work has provided evidence that it is possible to reliably track eye gaze in moving participants (see Hutton, 2019, as well as Steptoe et al., 2008, for the use of mobile eye-tracking in projection-based VR settings). The current study confirms that there is no methodological requirement for participants to remain static in virtual reality eye-tracking studies, also in the language sciences. The quality of the eye-tracking data was of such a nature that it allowed for a reliable analysis as only a relatively small number of trials per participant required rejection from the analysis due to technical limitations of the experimental apparatus. This observation opens up a wide variety of possibilities for future studies to study how people perceive the world while both their environment and their own behavior is dynamic and potentially interactive.

In general, the current study provides a methodological answer to recent theoretical calls for experimental research to develop more ecologically valid paradigms in the study of the human mind. Nevertheless, it has been argued that the term ecological validity might often be "ill-defined" and thus lacking specificity when it comes to the real-world phenomenon that ought to be explained or studied (Holleman et al., 2020). In line with this observation, we have clarified the context of the behaviors studied: Motion events can either be experienced passively (e.g., through being told about them) or actively, through being in motion. Both are relevant to the broader and comprehensive study of event cognition, yet the direct experience of motion events was previously difficult to test under the highly controlled circumstances required for most laboratory studies in this domain. The study of motion event perception may hence be best studied through either video stimuli or immersive VR as a function of whether the researcher is theoretically interested in the passive versus active experience of such events respectively. Matching aspects of the experimental setup with properties of the real-world phenomenon one is theoretically interested in reliably enhances the generalizability of experimental findings and its theoretical consequences to aspects of everyday behavior.

In addition, there has also been a recent call for more real-life neuroscience, which both embraces social-interactional and context-dependent aspects of human cognition, including using active participants (e.g., Schilbach et al., 2013; Shamay-Tsoory and Mendelsohn, 2019). Whilst the term "real-life" needs to be specified for the domain of interest, our findings suggest that data collection of participants in motion is both attainable and–depending on the research question–even desirable. To this end, our study adds to recent endeavors that suggest that also measures of brain activity (such as recorded through EEG or fNIRS) can be reliably obtained in dynamic participants (e.g., Askamp and van Putten, 2014; Gramann et al., 2014; Pinti et al., 2020), thereby opening up the (methodological) possibilities of combining dynamic VR with other measurement techniques.

Nevertheless, we encountered a number of technological obstacles and intrinsic limitations that will require further optimization for future studies.

First, in terms of ecological validity, our setup was restricted to participants only being able to walk on a straight (virtual) path. Additionally, this may have made the distinction between landmark and endpoint objects less evident to our participants. However, following existing behavioral studies we based our setup on (e.g., Flecken et al., 2014; Flecken et al., 2015), we opted to not instruct participants about this distinction. In line with our aim to create an environment mimicking the daily experience of moving and experiencing the world, we considered it good practice to avoid explicit explanation of the experimental manipulation. Technological advances in the development and use of omnidirectional treadmills may offer additional opportunities in these respects (e.g., Campos et al., 2012), and research labs may here also benefit from developments in the gaming industry. Relying solely on the VR data limits the extent to which ecological validity can be assessed. Follow-up work could aim to set up a study using eye tracking in the real world in order to obtain comparable data, though issues around experimental control will be introduced. On a more technical note, we lost some of our data through freezing of the eye-tracker and duplicates in the time stamps (see above). Future developments in VR-compatible eye-tracking hardware with accurate higher frame rates will likely circumvent such problems and reduce unnecessary data loss.

Second, in terms of data analysis, the current paradigm raised a number of challenges. As explained above, growth curve analysis is highly suitable for continuous eye-tracking data and has been used pre-dominantly in analysis of data elicited using the visual world paradigm (VWP, see Mirman, 2016). Similar to our setup, the VWP is typically used to study gaze behavior to competing stimulus items in the participants' field of view (e.g., Altmann and Kamide, 1999). Unlike in the conventional VWP, however, trials in our study were long and not structured by, for example, a target word to define time-windows. As a result, we were unable to base the choice of the length of time-windows and the number of time terms used in the analysis on well-established criteria. This observation confirms that, for a given topic under study, novel paradigms like the current one may require initially exploratory and subsequent hypothesis-driven studies to go hand in hand.

Third, we note that our study did not replicate previous differences between native German and native English speakers in the proportion of looks to endpoint objects. In contrast to previous studies in this field, German speakers in the present study behaved more "English-like:" they showed overall less bias in their looks towards endpoints. Despite testing our German sample group in their native language, the majority have reported also being proficient in English and/or Dutch. Depending on the proficiency and use of other languages (here: Dutch/English used by German speakers), a language-based endpoint bias may have been diminished in our study, meaning their everyday immersion in a second/third language environment could have influenced their behavior and thinking patterns. More research is needed to study the effects of L2/L3 proficiency and immersion, plus language of operation, in this specific experimental setting in more detail. Sampling was based on the availability of native speakers of both languages, as well as on the CAVE setup which is stationary and thus cannot be easily moved from one language community to another. Future work may investigate the use of head-mounted displays (HMDs) to alleviate the issue of location and allow for more flexible sampling, especially now HMDs also allow for. The advantage of using a CAVE, however, is that participants were able to see their own moving body while they walked across the various virtual environments, and not a customised virtual rendition of it as would have been necessary if a HMD had been used.

Furthermore, it could be the case that participants overall engaged with the 3-D stimuli to an extent that could have washed out any effects. After all, this was the first time for all participants that they moved around in a virtual environment. In line with this, our participants reported feeling immersed in the setup, and enjoyed moving through the virtual environments. This could suggest that, when participants are immersed in a setting mimicking real world dynamics, language effects on the perception of motion events may diminish or disappear. Differences between the sampling groups were observed in Phase 3, where English native speakers showed a stronger increase in looks towards the endpoint compared to German speakers. This unexpected result may at first sight resemble a goal bias for English speakers. Note however, that across all tasks, no strong goal bias was observed for either language group. However, both our two main behavioural tasks, which both relied on language and have previously shown robust and replicable effects (Athanasopoulos and Bylund, 2013; Flecken et al., 2014; Flecken et al., 2015), led to no observed differences between the language groups. The event description task explicitly asked participants to verbalize what they were observing, and the event similarity judgment task required participants to hold the video events in mind, likely utilizing language to do so (cf. Athanasopoulos and Bylund, 2013). Both behavioural tasks thus showed a similar pattern of results as the main eye-tracking virtual reality experiment.

In sum, we conclude that the VR setup implemented and tested in the current study is fit to be used for future experimental research in the field of psychology and beyond. Despite the discussed limitations, the novel setup enabled us to track looking behavior and its development over time (i.e., as participants were moving through space). This was something not possible with traditional video setups, as participants in such studies had less freedom in looking where they desired, but were instead restricted to the scene as it was displayed in the videos on a small computer screen. From a methodological viewpoint, the immersive VR setting with participants in motion provides a basis for future research by providing a less passive experimental approach, as for example advocated for by Hari et al. (2015). The current study is a successful attempt in utilizing the CAVE setup in combination with participants in motion, allowing for a first-person immersive experience, and thus potentially more ecologically valid testing of language effects on cognition.

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: https://hdl.handle.net/1839/d5507a85-a29f-42ae-9e58-2cc532d014e9.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The Ethics Committee of the Social Sciences Faculty of Radboud University, Nijmegen, Netherlands. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

ACKNOWLEDGMENTS

We thank Albert Russel and Jeroen Derks for their technical support, Eleanor Huizeling for her helpful discussion of the data, and Sebastian Sauppe for his input concerning growth curve analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frvir.2022.697934/full#supplementary-material

REFERENCES

Altmann, G. T. M., and Kamide, Y. (1999). Incremental Interpretation at Verbs: Restricting the Domain of Subsequent Reference. *Cognition* 73 (3), 247–264. doi:10.1016/s0010-0277(99)00059-1

Askamp, J., and van Putten, M. J. A. M. (2014). Mobile EEG in Epilepsy. *Int. J. Psychophysiol.* 91 (1), 30–35. doi:10.1016/j.ijpsycho.2013.09.002

Athanasopoulos, P., and Bylund, E. (2013). Does Grammatical Aspect Affect Motion Event Cognition? A Cross-Linguistic Comparison of English and Swedish Speakers. *Cogn. Sci.* 37 (2), 286–309. doi:10.1111/cogs.12006

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using Lme4. *J. Stat. Softw.* 67 (1), 1–48. doi:10.18637/jss.v067.i01

Blanco-Elorrieta, E., and Pylkkänen, L. (2018). Ecological Validity in Bilingualism Research and the Bilingual Advantage. *Trends Cognitive Sci.* 22 (12), 1117–1126. doi:10.1016/j.tics.2018.10.001

Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., and Bailenson, J. N. (2002). Immersive Virtual Environment Technology as a Methodological Tool for Social Psychology. *Psychol. Inq.* 13 (2), 103–124. doi:10.1207/s15327965pli1302_01

Bohil, C. J., Alicea, B., and Biocca, F. A. (2011). Virtual Reality in Neuroscience Research and Therapy. *Nat. Rev. Neurosci.* 12 (12), 752–762. doi:10.1038/nrn3122

Campos, J. L., Butler, J. S., and Bülthoff, H. H. (2012). Multisensory Integration in the Estimation of Walked Distances. *Exp. Brain Res.* 218 (4), 551–565. doi:10.1007/s00221-012-3048-1

Cañigueral, R., and Hamilton, A. F. D. C. (2019). The Role of Eye Gaze during Natural Social Interactions in Typical and Autistic People. *Front. Psychol.* 10, 560. doi:10.3389/fpsyg.2019.00560

Cruz-Neira, C., Leigh, J., Papka, M., Barnes, C., Cohen, S. M., Das, S., et al. (1993). "Scientists in Wonderland: A Report on Visualization Applications in the CAVE Virtual Reality Environment," in Proceedings of 1993 IEEE Research Properties in Virtual Reality Symposium, San Jose, CA, October 25–26, 1993 (IEEE), 59–66.

Eichert, N., Peeters, D., and Hagoort, P. (2018). Language-Driven Anticipatory Eye Movements in Virtual Reality. *Behav. Res.* 50 (3), 1102–1115. doi:10.3758/s13428-017-0929-z

Flecken, M., Gerwien, J., Carroll, M., and Stutterheim, C. V. (2015). Analyzing Gaze Allocation during Language Planning: A Cross-Linguistic Study on Dynamic Events1. *Lang. Cogn.* 7 (1), 138–166. doi:10.1017/langcog.2014.20

Flecken, M., von Stutterheim, C., and Carroll, M. (2014). Grammatical Aspect Influences Motion Event Perception: Findings from a Cross-Linguistic Non-Verbal Recognition Task. *Lang. Cogn.* 6 (1), 45–78. doi:10.1017/langcog.2013.2

Gramann, K., Ferris, D. P., Gwin, J., and Makeig, S. (2014). Imaging Natural Cognition in Action. *Int. J. Psychophysiol.* 91 (1), 22–29. doi:10.1016/j.ijpsycho.2013.09.003

Hari, R., Henriksson, L., Malinen, S., and Parkkonen, L. (2015). Centrality of Social Interaction in Human Brain Function. *Neuron* 88 (1), 181–193. doi:10.1016/j.neuron.2015.09.022

Heyselaar, E., Peeters, D., and Hagoort, P. (2020). Do we Predict Upcoming Speech Content in Naturalistic Environments? *Lang. Cogn. Neurosci.* 36, 440–461. doi:10.1080/23273798.2020.1859568

Heyselaar, E., Segaert, K., Wester, A. J., Kessels, R. P., and Hagoort, P. (2015). "Syntactic Operations Rely on Implicit Memory: Evidence from Patients with Amnesia," in The Individual Differences in Language Processing Across the adult Life Span Workshop, Nijmegen, Netherlands, December 10–11, 2015.

Holleman, G. A., Hooge, I. T. C., Kemner, C., and Hessels, R. S. (2020). The 'Real-World Approach' and its Problems: A Critique of the Term Ecological Validity. *Front. Psychol.* 11, 721. doi:10.3389/fpsyg.2020.00721

Hömke, P., Holler, J., and Levinson, S. C. (2018). Eye Blinks are Perceived as Communicative Signals in Human Face-To-Face Interaction. *PLoS One* 13 (12), e0208030. doi:10.1371/journal.pone.0208030

Hutton, S. B. (2019). "Eye Tracking Methodology," in *Eye Movement Research. Studies in Neuroscience, Psychology and Behavioral Economics*. Editors C. Klein and U. Ettinger (Cham: Springer), 207–308. doi:10.1007/978-3-030-20085-5_8

Klein, W. (1994). *Time in Language*. Routledge, London: Psychology Press.

Knoeferle, P. (2015). "Visually Situated Language Comprehension in Children and in Adults," in *Attention and Vision in Language Processing*. Editors R. Mishra, N. Srinivasan, and F. Huettig (New Delhi: Springer), 57–75. doi:10.1007/978-81-322-2443-3_4

Krohn, S., Tromp, J., Quinque, E. M., Belger, J., Klotzsche, F., Rekers, S., Chojecki, P., de Mooij, J., Akbal, M., McCall, C., Villringer, A., Gaebler, M., Finke, C., and Thöne-Otto, A. (2020). Multidimensional Evaluation of Virtual Reality Paradigms in Clinical Neuropsychology: Application of the VR-Check Framework. *J. Med. Internet Res.* 22 (4), e16724. doi:10.2196/16724

Magliano, J. P., Radvansky, G. A., Forsythe, J. C., and Copeland, D. E. (2014). Event Segmentation during First-Person Continuous Events. *J. Cognitive Psychol.* 26 (6), 649–661. doi:10.1080/20445911.2014.930042

Mirault, J., Guerre-Genton, A., Dufau, S., and Grainger, J. (2020). Using Virtual Reality to Study Reading: An Eye-Tracking Investigation of Transposed-Word Effects. *Methods Psychol.* 3, 100029. doi:10.1016/j.metip.2020.100029

Mirman, D. (2016). *Growth Curve Analysis and Visualization Using R*. New York, NY: CRC Press.

Nölle, J., Kirby, S., Culbertson, J., and Smith, K. (2020). "Does Environment Shape Spatial Language? A Virtual Reality Experiment," in The Evolution of Language: Proceedings of the 13th International Conference (EvoLang13), Brussels, Belgium, April 14-17, 2020. doi:10.17617/2.3190925

Pan, X., and Hamilton, A. F. D. C. (2018). Why and How to Use Virtual Reality to Study Human Social Interaction: The Challenges of Exploring a New Research Landscape. *Br. J. Psychol.* 109 (3), 395–417. doi:10.1111/bjop.12290

Parsons, T. D. (2015). Virtual Reality for Enhanced Ecological Validity and Experimental Control in the Clinical, Affective and Social Neurosciences. *Front. Hum. Neurosci.* 9, 660. doi:10.3389/fnhum.2015.00660

Peelle, J. E., and Van Engen, K. J. (2020). Time Stand Still: Effects of Temporal Window Selection on Eye Tracking Analysis. Preprint. doi:10.31234/osf.io/pc3da

Peeters, D. (2018). A Standardized Set of 3-D Objects for Virtual Reality Research and Applications. *Behav. Res.* 50 (3), 1047–1054. doi:10.3758/s13428-017-0925-3

Peeters, D., and Dijkstra, T. (2018). Sustained Inhibition of the Native Language in Bilingual Language Production: A Virtual Reality Approach. *Bilingualism* 21 (5), 1035–1061. doi:10.1017/s1366728917000396

Peeters, D. (2019). Virtual Reality: A Game-Changing Method for the Language Sciences. *Psychon. Bull. Rev.* 26 (3), 894–900. doi:10.3758/s13423-019-01571-3

Pinti, P., Tachtsidis, I., Hamilton, A., Hirsch, J., Aichelburg, C., Gilbert, S., et al. (2020). The Present and Future Use of Functional Near-Infrared Spectroscopy (fNIRS) for Cognitive Neuroscience. *Ann. N. Y. Acad. Sci.* 1464 (1), 5–29. doi:10.1111/nyas.13948

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Sassenhagen, J., and Alday, P. M. (2016). A Common Misapplication of Statistical Inference: Nuisance Control with Null-Hypothesis Significance Tests. *Brain Lang.* 162, 42–45. doi:10.1016/j.bandl.2016.08.001

Sauppe, S. (2017). Word Order and Voice Influence the Timing of Verb Planning in German Sentence Production. *Front. Psychol.* 8, 1648. doi:10.3389/fpsyg.2017.01648

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., et al. (2013). Toward a Second-Person Neuroscience. *Behav. Brain Sci.* 36 (4), 393–414. doi:10.1017/s0140525x12000660

Shamay-Tsoory, S. G., and Mendelsohn, A. (2019). Real-Life Neuroscience: An Ecological Approach to Brain and Behavior Research. *Perspect. Psychol. Sci.* 14 (5), 841–859. doi:10.1177/1745691619856350

Steptoe, W., Wolff, R., Murgia, A., Guimaraes, E., Rae, J., Sharkey, P., et al. (2008). "Eye- Tracking for Avatar Eye-Gaze and Interactional Analysis in Immersive Collaborative Virtual Environments," in Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work, San Diego, CA, November 8–12, 2008. doi:10.1145/1460563.1460593

Swallow, K. M., Kemp, J. T., and Candan Simsek, A. (2018). The Role of Perspective in Event Segmentation. *Cognition* 177, 249–262. doi:10.1016/j.cognition.2018.04.019

Tromp, J., Peeters, D., Meyer, A. S., and Hagoort, P. (2018). The Combined Use of Virtual Reality and EEG to Study Language Processing in Naturalistic Environments. *Behav. Res.* 50 (2), 862–869. doi:10.3758/s13428-017-0911-9

Von Stutterheim, C., Andermann, M., Carroll, M., Flecken, M., and Schmiedtová, B. (2012). How Grammaticized Concepts Shape Event Conceptualization in Language Production: Insights from Linguistic Analysis, Eye Tracking Data, and Memory Performance. *Linguistics* 50 (4), 833–867. doi:10.1515/ling-2012-0026

Willems, R. M. (2015). *Cognitive Neuroscience of Natural Language Use*. Cambridge: Cambridge University Press.

Conflict of Interest: The authors declare 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.