Landscape and Urban Planning 189 (2019) 129–138

Contents lists available at ScienceDirect

Landscape and Urban Planning

journal homepage: www.elsevier.com/locate/landurbplan

Research Paper

The utilization of immersive virtual environments for the investigation of environmental preferences

Amit Birenboim[a,*], Martin Dijst[b], Dick Ettema[c], Joost de Kruijf[d], Geert de Leeuw[d,e], Nico Dogterom[f]

[a] *Department of Human Geography and Spatial Planning, Utrecht University and Department of Geography and the Human Environment, Tel Aviv University, Israel*
[b] *Department of Human Geography and Spatial Planning, Utrecht University and Department of Urban Development and Mobility, Luxembourg Institute of Socio-Economic Research, Luxembourg*
[c] *Department of Human Geography and Spatial Planning, Utrecht University, Netherlands*
[d] *Academy for Urban Development, Logistics & Mobility, Breda University of Applied Sciences, Netherlands*
[e] *XTNT Experts in Traffic and Transport, Netherlands*
[f] *Department of Human Geography and Spatial Planning, Utrecht University and Goudappel Coffeng BV, Netherlands*

ARTICLE INFO

Keywords:
Virtual reality
Immersive virtual environments
Cycling
Environmental preferences
Stated preference
Spatial presence

ABSTRACT

The article discusses the feasibility and benefits of using immersive virtual environments (IVEs) to gauge the environmental preferences of individuals. The discussion is based on the results of a stated preference conjoint experiment employed within an IVE. In the experiment, participants were asked to rate and rank their cycling experience during and after they had cycled a few virtual routes with changing environmental characteristics. Participants repeated the experiment a week later to allow the examination of the test–retest reliability of the method. Presence level—namely the extent to which one has an actual sense of being in the simulated world—was computed using the ITC–SOPI questionnaire. The scores were compared with an equivalent, more traditional, still images conjoint task that was administrated to a control group. Presence level was significantly higher in the IVE compared to the still images experiment. This finding supports the notion that IVEs may yield greater external validity due to their higher level of realism. Relatively low test–retest reliability scores between the two IVE experiment rounds were obtained. This might be explained by the participants' low familiarity with IVEs, which in the first round diverted their attention from the conjoint task itself. In contrast, the test–retest scores of post-IVE evaluations, which are considered more cognitive in their nature, were satisfactory. Implications of the experiments and suggestions for future research are discussed.

1. Introduction

Physical and social environments are thought to have a considerable impact on human daily behavior and lifestyle. Environmental factors have been found to affect, for example, the number and quality of social interactions (Kaźmierczak, 2013; Zhao, Dijst, & Chai, 2016), dietary behavior (Burgoine, Forouhi, Griffin, Wareham, & Monsivais, 2014; Poelman et al., 2018), commuting and transportation preferences (Cervero, 2002; Ghekiere et al., 2015), the decision to engage in physical activity (Jansen, Kamphuis, Pierik, Ettema, & Dijst, 2018), and mood, mental health, and wellbeing (Birenboim, 2018; Evans, 2003). Studies that investigate the effect of environmental factors on human behavior may take the form of either real-world, ecological studies in which participants' behavior is observed using various techniques (e.g., questionnaires and diaries, interviews, tracking technologies such as GPS), or more experimental, lab approaches. While the latter approaches are considered to have higher internal validity and therefore to be superior in terms of causality inference, the former approaches represent real-life situations and are therefore thought to have a greater external or ecological validity.

Stated preference (SP) is a common group of experimental methods used to evaluate environmental changes that cannot be tested in real-world observational approaches. In SP research, people are asked about their preferences in hypothetical situations. It usually involves a choice, ranking, or rating task, in which respondents have to evaluate a set of alternatives that differ on predefined characteristics of interest (Louvière & Timmermans, 1990; Louvière, Hensher, & Swait, 2000). The SP approach is commonly applied in fields such as marketing,

---

* Corresponding author at: Department of Geography and the Human Environment, Tel Aviv University, Zelig 10, 6997801 Tel Aviv, Israel.
E-mail addresses: abirenboim@tauex.tau.ac.il (A. Birenboim), Martin.Dijst@liser.lu (M. Dijst), D.F.Ettema@uu.nl (D. Ettema), Kruijf.J@buas.nl (J. de Kruijf), Leeuw.G@buas.nl (G. de Leeuw), ndogterom@goudappel.nl (N. Dogterom).

https://doi.org/10.1016/j.landurbplan.2019.04.011
Received 18 November 2018; Received in revised form 1 March 2019; Accepted 12 April 2019
Available online 03 May 2019
0169-2046/

transportation, planning, and housing (Earnhart, 2002; Louvière, 1979; Molin, Oppewal, & Timmermans, 2001). Other major advantages of SP experiments are their low cost and relative ease of administration and analysis, compared to real-world, observational studies.

It is of critical importance that respondents in SP studies that focus on environmental changes have a clear and accurate impression of the alternative environments and situations presented to them (Orzechowski, Arentze, Borgers, & Timmermans, 2005). Traditionally, most SP research involved verbal descriptions of alternatives and their attributes. This often led to SP techniques being criticized for relying heavily on respondents' imagination, possibly leading to considerable error variance due to poor evaluability (Bateman, Day, Jones, & Jude, 2009; Farooq, Cherchi, & Sobhani, 2018). Being aware that textual representations cannot always adequately convey the nature and complexity of certain decision contexts, studies have started to incorporate visualization tools such as GIS-based assistance, images, and photographs to enhance task realism and avoid biased responses (Caulfield, Brick, & McCarthy, 2012; Tilahun, Levinson, & Krizek, 2007; Verhoeven et al., 2017; Yamada & Thill, 2003). However, these tools provide only a static and often rudimentary impression of the environment the researcher intends to present (Jansen, Boumeester, Coolen, Goetgeluk, & Molin, 2009). We hypothesized that a study involving a more virtual dynamic interaction with the environment, such as in the case of virtual movement through space, would facilitate more realistic responses of participants. Moreover, static, mono-sensory (focusing only on the visual) simulations of environmental changes are not likely to produce a real embodied experience that is assumed to significantly impact human responses through multi-sensory sensations (Patterson, Darbani, Rezaei, Zacharias, & Yazdizadeh, 2017); these include the vestibular and proprioception senses, which are highly relevant to interaction in virtual environments (Sanchez-Vives & Slater, 2005).

In this paper, we discuss the implementation of immersive virtual reality technology in an SP experiment that investigated the environmental preferences of cyclists. Dutch participants cycled through various virtual environments in a simulator based on a head-mounted display, after which they were asked to rate and rank their cycling experience. Much research has been carried out on the associations between the built environment and cycling (Fraser & Lock, 2011; Saelens, Sallis, & Frank, 2003; Van Cauwenberg et al., 2018; Wendel-Vos et al., 2004), and some of it utilized SP techniques (Ghekiere et al., 2015; Vedel, Jacobsen, & Skov-Petersen, 2017). However, since cycling involves a direct and dynamic interaction with the surrounding environment, it was hypothesized that the incorporation of immersive virtual reality techniques would result in greater external validity compared to traditional SP studies. O'Hern, Oxley, and Stevenson (2017) have already established the ability to produce real-life equivalent cycling behaviors in a head-mounted display bicycle simulator. These behaviors included cycling speed, head movements, lane position, and relative distance to cars. However, the study did not deal with aspects of subjective experiences and did not utilize SP techniques. Similarly, Nazemi, van Eggermond, Erath, Schaffner, and Axhausen (2018) assessed the perception of relative speed changes of cars and lanes width in a cycling simulator. They found that participants could detect relative changes in cars speed and lanes width in virtual environments except in cases where changes were minor (changes of < 10 km per hour in speed, and < 30 cm in lane width).

The present research aims at evaluating the qualities and methodological advantages and weaknesses that are associated with the implementation of SP experiments within immersive virtual environments (IVEs) for studying environmental preferences in general and cycling behavior more specifically. To do so, the study first tests the hypothesis that IVEs generate greater realism by comparing the level of realism in a cycling IVEs with equivalent still images environments. The study then turns to evaluate the test–retest reliability of SP experiments within IVEs. Test-retest procedure examines whether a measurement tool can reproduce the same result given the same conditions in, at least, two different time periods. It is an important indicator of the reliability of the tool and hence its validity. The feasibility, advantages and disadvantages of implementing SP experiments within IVEs are discussed in the concluding sections based on the literature and our experience with the current study.

2. Virtual reality and immersive virtual environments

In the late 1990 s, researchers began to show more interest in the utilization of immersive virtual technology as a scientific research tool. The main advantage associated with the technology lies in its potential to address the long-standing trade-off problem between mundane realism and experimental control that is encountered in many experiments on human perceptions and behaviors (Blascovich et al., 2002; Fox, Arena, & Bailenson, 2009; Loomis, Blascovich, & Beall, 1999; Ruotolo et al., 2013; Smith, 2015). On the one hand, the experimental lab design gives researchers full control over the stimuli to which participants are exposed, and it allows them to rule out possible impacts of exogenous confounders that might be present in the field. These are two essential prerequisites to be able to identify true (causal) relationships. Real-life contexts seldom allow the experimental control that is critical for the isolation of the effects of environmental change or interventions (Handy, van Wee, & Kroesen, 2014; Krizek, Handy, & Forsyth, 2009). On the other hand, in laboratory research, the ecological validity—-namely the extent to which the research task approximates a situation as experienced in real life—is generally reduced, thus lowering the likelihood that the manipulation effect that was enacted in the experiment will also work in real life (Blascovich et al., 2002).

For this reason, new technologies have been enthusiastically embraced by the research community in the attempt to decrease the degree of trade-off between realism and control in laboratory experiments on human behavior and perceptions (Loomis et al., 1999; Patterson et al., 2017; Sylcott, Orsborn, & Cagan, 2016). Video and audio recordings have proven to be valuable supplements to the traditional word and picture-based representations of situations of interest. More recently, computer-based visualizations of environments have been increasingly used in research, and have sometimes even enabled interaction between the participant and the virtual objects and/or environment depicted (Borgers, Brouwer, Kunen, Jessurun Joran, & Janssen Ingrid, 2010; Dijkstra, van Leeuwen, & Timmermans, 2003; Kahlert & Schlicht, 2015). These technologies are often called virtual reality (VR) in the literature as they increase realism and participant engagement. However, most of these technologies still provide a VR experience that is characterized by a discontinuity in temporal and spatial realities, with an evident perceptual gap between the physical world in which the participant is situated and the non-physical world that is depicted in a 2D interface on a distant screen.

Immersive virtual environment (IVE) technologies seek to reproduce reality in a more convincing way by offering participants a virtual sensation that is both more inclusive and perceptually richer. The simulated environments typically completely surround the participant through the use of, for example, VR glasses (head-mounted display) or a virtual display on the walls around the participants in a room-sized cube (cave automatic virtual environment; CAVE), and a dynamic display enables a direct coupling between the participant's motor actions and the simulation. IVEs are capable of delivering visual (sight), auditory (hearing), haptic (touch), olfactory (smell), and gustatory (taste) sensations to the participant's senses, as well stimulating the vestibular (balance) and proprioception (relative position) senses. However, the presentation of stimuli other than visual and auditory stimuli remains a big technological challenge (Smith, 2015). IVE technologies have been especially welcomed in places where field experiments are impossible, dangerous, or expensive, such as in the investigation of human behavior during evacuations in tunnels (Cosma, Ronchi, & Nilsson, 2016; Ronchi et al., 2015), children's street-crossing behavior (Grechkin, Chihak, Cremer, Kearney, & Plumert, 2013; Morrongiello, Corbett, Milanovic, Pyne, & Vierich, 2015), and drivers' and pedestrians behavior (Farooq et al., 2018; Risto & Martens, 2014; Rumschlag et al., 2015). The growing availability and popularization of immersive technologies—most notably of affordable head-mounted display systems (e.g., Oculus Rift, HTC Vive, Sony PlayStation VR)—in very recent years, has increased the attractiveness of the implementation of IVE technology for scientific research purposes. These new display systems have complementary input devices, such as omnidirectional treadmills and hand-worn controllers that provide discrete input in the form of buttons and/or continuous input by top-mounted joysticks or touchpads (Anthes, Garcia-Hernandez, Wiedemann, & Kranzlmüller, 2016, p. 3). All these devices allow a more realistic interaction with the virtual environment, creating a greater sense of virtual realism.

To evaluate the extent to which IVE technology is able to faithfully represent a real-life environmental experience, psychological research on the experiences of virtual environments introduced the concepts of "immersion" and "presence." Immersion describes the display of information according to the number of senses that are addressed, the richness of the information that is mediated, and the degree to which the participants find themselves included by and interacting with the displayed environment, in the absence of stimuli from the outside world (Slater & Wilbur, 1997). While immersion is an objective measure of the display of information by a technology, the concept of presence is about the concomitant human response, that is, the extent to which the participant has a sense of being in the simulated world (Gaggioli, Bassi, & Delle Fave, 2003; Slater & Wilbur, 1997). It can be defined as "the subjective experience of being in one place or environment, even when one is physically situated in another… As applied to a virtual environment, presence refers to experiencing the computer-generated environment rather than the actual physical locale" (Witmer & Singer, 1998, p. 225). It is therefore closely related to sense of embodiment (Kilteni, Groten, & Slater, 2012). Immersion and presence are thus related concepts and can be seen as correlates; that is, the more immersive a VR technology, the more participants believe that they are physically present in the simulated environment (de Kort, Meijnders, Sponselee, & IJsselsteijn, 2006; Gorini, Capideville, De Leo, Mantovani, & Riva, 2011; Slater & Wilbur, 1997; Waller, Beall, & Loomis, 2004). Presence is known to have a moderate positive impact on task performance (Nash, Edwards, Thompson, & Barfield, 2000; Witmer & Singer, 1998). More important to our case is that presence is assumed to enhance the "ecological qualities" of virtual environments. In other words, it is expected that the greater the sense of immersion and presence in a virtual environment is and thus the sense of realism, the more consistent the behaviors in this environment will be with everyday behaviors and responses (Slater & Wilbur, 1997). Following this assumption, greater ecological validity is expected when people experience a high level of embodiment or presence in the IVE.

As explained above, it is also important to understand the impact on SP tasks of experiences that are produced within IVEs. Although the SP technique is an established research method and stated preferences have generally been shown to be temporally stable (Bryan, Gold, Sheldon, & Buxton, 2000; Louvière et al., 2000), most of this research relies on surveys that describe objects and situations based on texts or images. As IVE presents participants with an experience that is quite different in scope and nature, it is critical to assess the influence of the application of IVE technology on the consistency of preferences before drawing any explicit conclusions (Farooq et al., 2018; Patterson et al., 2017).

3. Methods

3.1. Conjoint experiment within IVE

To test the feasibility and benefits of implementing VR technology for the investigation of environmental preferences, the present study utilized a conjoint experiment within IVE. A conjoint experiment is an SP research technique that is used to assess the value people assign to specific attributes of products (which may also include physical environments) or services by systematically varying the levels of the characteristics (or attributes) of the product or situation under investigation. This allows researchers to estimate the relative effects of each of these attributes. Conjoint experiments are widely used in marketing, but are often also implemented to assess transportation (Ghekiere et al., 2015), planning (Katoshevski & Timmermans, 2001; Orzechowski et al., 2005), and other environment-related preferences (Alriksson & Öberg, 2008). The conjoint experiment was administrated twice in order to assess the reliability of the method.

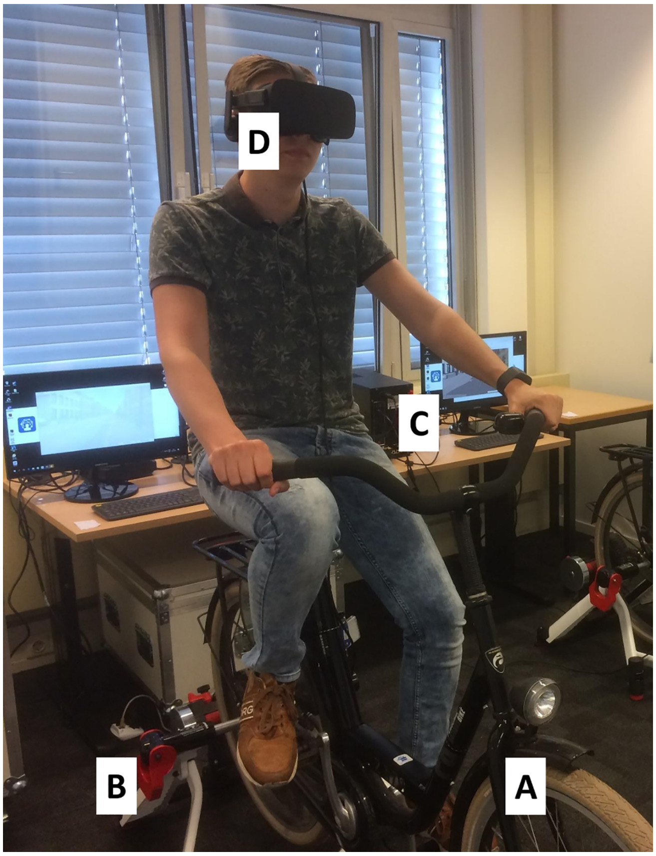

Fig. 1. The physical setup of the VR experiments: (A) bicycle, (B) electromagnetic trainer (Elite RealAxiom Wired), (C) computer, (D) head-mounted display (Oculus Rift CV1).

3.1.1. Materials

Fig. 1 presents the physical setup of the experiment. The hardware included a standard Dutch bicycle (Fig. 1-A) that was affixed to an electromagnetic trainer (Elite RealAxiom Wired; Fig. 1-B). The bicycle trainer, which was connected to a computer (Fig. 1-C), transmitted data from which the virtual cycling movement parameters could be extracted. An Oculus Rift CV1 headset (Fig. 1-D) was connected to the same computer and displayed the graphics of the IVE (see Fig. 2 for a static illustration). Participants could accelerate and brake in a natural way using the bicycle pedals. However, due to technical constraints, steering was not possible (i.e., participants could only cycle forward in a straight line).

The IVE, which was developed by a third-party company, was designed as a typical Dutch street in which several predefined attributes could be modularly modified. In the present study, three factors were manipulated using binary (two-level) attributes: 1) The factor

Fig. 2. Still images that were extracted from the experiment's IVE (four of the eight environments). The text about the environmental factors was displayed only in the post-IVE ranking task and in the still images experiment.

"greenness" had a green (Fig. 2-A and 2-C) and a no green (Fig. 2-B and 2-D) attribute; 2) the factor "bicycle path" had a wide and well-separated bicycle lane attribute (Fig. 2-A and 2-B) and a narrow and poorly-separated bicycle lane attribute (Fig. 2-C and 2-D); and 3) the factor "traffic volume" could be assigned either a low volume of pedestrians and cyclists (Fig. 2-A and 2-D) or a high volume of pedestrians and cyclists. The maximum number of environments (i.e., combinations of attributes) that could be generated from this setup is eight (Fig. 2 presents four of the eight combinations). A full factorial conjoint experiment that included all eight environmental combinations was used in the experiments.

3.1.2. *Procedure*

The participants arrived independently at the VR lab for their scheduled appointments. The aims and procedure of the experiment were explained, as were the possible annoyances that the usage of head-mounted displays may involve (e.g., dizziness). After signing an informed consent form, they were asked to seat themselves comfortably on the bikes (see Fig. 1). The research assistant then asked them to put the Oculus Rift headset on their head. If necessary, the assistant adjusted the fit of the headset.

The experiment started with a 300-meter introductory cycling segment, which was the same for all participants. In this segment, participants were instructed to familiarize themselves with the IVE, namely to accelerate, brake, and look around. This was followed by four cycling segments that were similar in length and assigned randomly to each participant. In each segment, a different combination of the three factors that were investigated (greenness, bicycle path, and traffic volume) was introduced. At the end of each segment, participants rated how aesthetically pleasing, how safe, and how enjoyable the environment the cycling experience had been on a 7-point Likert scale. Rating was conducted within the virtual environment by moving the head and staring at the desired answer. In order to avoid long exposure to the IVE, which may increase the chance of negative side effects, it was decided to limit the experiments to ~10 min. For this reason, each participant experienced and rated only five cycling segments (i.e., an introduction and four trials). Since a full factorial design included eight profiles, it took two participants to complete a full conjoint trial.

At the end of the cycling session, the participants were instructed to self-complete a questionnaire on a laptop that was located in an adjacent room. In the first part of the questionnaire, they were asked to rank the cycling segments they had just experienced from the most attractive to the least attractive. In order to ensure that participants remembered the order and the attributes of the environments in which they had just cycled, still images of the environments (taken from the IVE) were displayed within the questionnaire. It is assumed that this more traditional type of questionnaire rely mainly on cognitive assessment, which reflects prototypical beliefs and preferences often culturally appropriated compared to VR experiments. The second part of the questionnaire included the ITC–Sense of Presence Inventory (ITC–SOPI), which is commonly used for assessing the sense of presence (Lessiter, Freeman, Keogh, & Davidoff, 2001). The ITC–SOPI includes four subscales that represent four independent dimensions of presence, namely sense of physical space (spatial presence), engagement, ecological validity (naturalness), and negative effects (e.g., nausea). It was designed as a cross-media presence measure that can be implemented in various platforms and in different environmental contexts.

In the third and final part of the questionnaire, participants were asked to supply their basic personal details, including age, gender, and their previous experience with VR. After completing the survey, an appointment for the follow-up session was scheduled for the coming week. The follow-up sessions were identical to the first sessions, except the ITC-SOPI and the personal detail questionnaires were not administrated again. The follow-up sessions were implemented to test repeatability (i.e., the level at which participants could reproduce their original scores), which is commonly known as test–retest reliability.

3.1.3. *Participants*

Data collection took place over four nonconsecutive days in June 2017. The participants were students and staff members from Breda University of Applied Sciences, in Breda, the Netherlands (see Table 1 for sample characteristics). Students were approached during their classes (after consulting the courses' lecturers) and through posters and flyers that were distributed throughout the campus. Breda University staff members were recruited through emails or by a research assistant who approached them directly. As an incentive, ten 20-euro vouchers were given as prizes in a lottery among participants who completed the entire experiment. The group of participants who took part in the first experiment comprised 86 individuals with a mean age of 28.2 (std 12.4); 59 (68.6%) were males and 61 (70.9%) were students. The group of participants in the second round comprised 66 participants (76.7% returned for the retest session) with a mean age of 28.9 (std 13.4); 45

Table 1
Samples' characteristics.

	Still images experiment	VR test experiment			VR retest experiment		
	all (students)	students	staff	All	students	staff	all
Frequency	55	61	25	86	46	20	66
Mean age (std)	21.38 (1.95)	21.1 (2.5)	45.2 (9.9)	28.2 (12.4)	20.9 (2.2)	47.4 (9.5)	28.9 (13.4)
Gender (male)	70.9%	72.9%	60.0%	68.6%	71.7%	60.0%	68.2%
Nationality (Dutch)	92.7%	67.2%	96.0%	75.6%	71.7%	95.0%	78.8%
VR experienced	n/a	18.0%	32.0%	22.1%	15.2%	30.0%	19.7%

Table 2
ITC–SOPI scores of the still images and the VR experiment.

	Still images experiment		VR test experiment						
	All		All		Experienced*		Non-experienced*		Experienced vs. non-experienced
Scale	Mean	Cronbach's α	Mean	Cronbach's α	Mean	Cronbach's α	Mean	Cronbach's α	*t*-test (1-tailed)
Spatial presence (19)	2.712	0.951	3.061	0.891	2.864	0.870	3.117	0.894	t = 1.777, p = 0.04
Ecological validity (5)	n/a	n/a	3.449	0.717	3.158	0.804	3.531	0.643	t = 1.875, p = 0.037
Engagement (13)	n/a	n/a	3.606	0.759	3.368	0.613	3.673	0.762	t = 2.767, p = 0.004
Negative effects (6)	n/a	n/a	1.901	0.888	1.982	0.941	1.878	0.860	t = -0.473, p = 0.318
	n = 53		n = 86		n = 19		n = 67		

– Number in parentheses = number of items used for each scale.
– T-test between spatial presence of still images and test round (all participants) $t = -3.2$, $p < 0.001$.
* Experienced participants had used VR at least several times in the past. Non-experienced participants had never used VR or had used it only a few times in the past.

(68.2%) were males and 46 (69.7%) were students.

3.1.4. Analysis

The average scores of the four independent ITC–SOPI subscales were calculated, as was a Cronbach's α to examine the internal consistency of the items comprising each subscale. A comparison was made between VR-experienced and non-experienced individuals, since it was anticipated that experienced participants who had encountered IVEs several times before would be more sensitive to imperfect graphical representations and less overwhelmed by the VR experience.

The test–retest reliability of the IVE experiment items aesthetics, safeness, and enjoyment were assessed using the ICC (3,1) and ICC (3,k) variant of the intraclass correlation coefficient (Weir, 2005). The reliability of the post-IVE ranking task (ordinal scale) was assessed using a weighted kappa.

Part-worth utilities of the attributes in the conjoint experiments were assessed using a linear regression where variables' coefficient serve as the part-worth function (Orme, 2010; Rao, 2014). The relative importance of each factor (i.e., the weight that the factor had in the decision process) was evaluated by calculating the range of the part-worths of each factor (difference between highest and lowest coefficients) divided by the sum of the ranges across all the factors (Orme, 2010, Chapter 9).

3.2. Traditional still images conjoint experiment

A more traditional online still images conjoint experiment was administrated to 55 students from Breda University who had not participated in the original VR experiment. The sample's characteristics are presented in Table 1. The conjoint questionnaire was similar in design to that used in the IVE experiment described above. In the first part, participants were asked to rate on 7-point Likert scale (1) the aesthetics and safety of the cycling environment in the images presented to them and (2) how much they would enjoy cycling in that environment. The images of the cycling environment had been extracted from the IVE experiment and included text that described the attributes of the environment (see Fig. 2). It was decided to include text in the images since it is often the case that conjoint experiments that employ still images include also a textual description. In the second part of the questionnaire, participants had to rank the cycling segments that were presented to them, just like the participants in the post-evaluation questionnaire of the IVE experiment. The third part included the “spatial presence” subscale of the ITC–SOPI. It was decided to administrate only this subscale since (1) it is arguably the most common and relevant dimension of presence in relation to environmental preferences, and (2) to keep the questionnaire relatively short so that the participants could complete it while giving it their full attention. In the fourth and final part of the experiment, the participants completed a personal details questionnaire. The still images conjoint experiment was used to evaluate the differences between the standard and the IVE form of SP experiments.

4. Results

4.1. Presence

The ITC–SOPI presence scores are presented in Table 2. ITC–SOPI scales are normally used as a relative indicator that allows a comparison of the presence levels between different mediums (e.g., immersive vs. non-immersive display) or conditions. The spatial presence score was significantly lower in the still images experiment compared to the scores reported in the first IVE experiment (2.712 vs. 3.061, $t = -3.2$, $p < 0.001$). This means that, as expected, the IVEs generated a greater sense of presence and a stronger embodied experience. Cronbach's αs of the subscales were high (0.951 and 0.891, respectively), indicating a very good internal consistency of the subscale items in both experiments.

VR-experienced participants reported lower levels of spatial presence in the IVE experiment compared to less experienced participants (2.864 vs. 3.117, $t = 1.777$, $p = 0.04$). As noted above, it is assumed that experienced VR users will be less overwhelmed by the immersiveness of the displayed environment and more sensitive to imperfect graphical representations, which may reduce their sense of presence. The lower sense of presence of VR-experienced users was also reflected

Table 3
Conjoint experiment results of experience rating of still images and IVE experiments (test and retest).

	Still images experiment				VR test experiment				VR retest experiment			
	Coefficients	t	p-value	Importance	Coefficients	t	p-value	Importance	Coefficients	t	p-value	Importance
A) Dependent variable: do you find the cycling environment aesthetically pleasing (on a 7-point scale)?												
(Constant)	3.388	19.916	< 0.001		4.573	30.253	< 0.001		4.357	33.331	< 0.001	
Green	1.775	10.302	< 0.001	77.5%	0.901	5.941	< 0.001	70.6%	1.507	11.437	< 0.001	90.1%
Wide	0.413	2.396	0.017	18.0%	0.202	1.332	0.184	15.8%	0.078	0.588	0.557	4.6%
Low traffic	−0.102	−0.594	0.553	4.5%	−0.172	−1.136	0.257	13.5%	−0.088	−0.671	0.503	5.3%
	R2(adjusted) = 0.334 (0.325)				R2(adjusted) = 0.101 (0.093)				R2(adjusted) = 0.343 (0.336)			
	N = 220 (55 respondents)				N = 344 (86 respondents)				N = 256 (64 respondents)			
					Test vs. retest comparison							
					ICC(3,1) = 0.583, ICC(3,k) = 0.736, wkappa = 0.407, r = 0.584(Pvalue < 0.001), hit rate = 36.6%							
B) Dependent variable: how much did you enjoy cycling in this environment? (on a 7-point scale)?												
(Constant)	3.742	23.884	< 0.001		4.712	34.523	< 0.001		4.755	37.351	< 0.001	
Green	1.362	8.587	< 0.001	66.5%	0.816	5.979	< 0.001	64.3%	0.941	7.328	< 0.001	88.8%
Wide	0.632	3.983	< 0.001	30.8%	0.356	2.607	0.01	28.0%	0.068	0.532	0.595	6.4%
Low traffic	−0.056	−0.352	0.725	2.7%	0.098	0.719	0.473	7.7%	0.051	0.395	0.693	4.8%
	R2(adjusted) = 0.280 (0.270)				R2(adjusted) = 0.116 (0.108)				R2(adjusted) = 0.183 (0.173)			
	N = 220 (55 respondents)				N = 337 (85 respondents)				N = 247 (62 respondents)			
					Test vs. retest comparison							
					ICC(3,1) = 0.504, ICC(3,k) = 0.670, wkappa = 0.349, r = 0.513(Pvalue < 0.001), hit rate = 35.5%							
C) Dependent variable: how safe did you feel cycling in this environment? (on a 7-point scale)?												
(Constant)	3.750	22.723	< 0.001		4.549	31.542	< 0.001		4.523	30.345	< 0.001	
Green	0.638	3.816	< 0.001	36.7%	0.219	1.511	0.132	16.4%	0.348	2.317	0.021	24.7%
Wide	0.997	5.965	< 0.001	57.4%	0.755	5.216	< 0.001	56.6%	0.694	4.619	< 0.001	49.2%
Low traffic	0.103	0.617	0.538	5.9%	0.359	2.483	0.014	27.0%	0.368	2.452	0.015	26.1%
	R2(adjusted) = 0.180 (0.169)				R2(adjusted) = 0.096 (0.088)				R2(adjusted) = 0.118 (0.107)			
	N = 220 (55 respondents)				N = 343 (86 respondents)				N = 255 (64 respondents)			
					Test vs. retest comparison							
					ICC(3,1) = 0.554, ICC(3,k) = 0.713, wkappa = 0.385, r = 0.556(Pvalue < 0.001), hit rate = 34.8%							

in the two other subscales (ecological validity and engagement), though it should be noted that the Cronbach's αs were relatively lower (e.g., 0.613 for the engagement scale for VR-experienced participants), indicating inferior consistency between the subscale items compared to the spatial presence subscale. The negative effects (e.g., disorientation) subscale was similar for both experienced and non-experienced participants. It is therefore assumed that the undesired side effects did not cause the difference in sense of presence.

4.2. Test–retest reliability of IVE conjoint experiments

4.2.1. Experience rating task

Table 3 presents the results of the conjoint experiments in which participants were asked to rate three dimensions of their cycling experience, namely A) aesthetics, B) enjoyment, and C) safety. The results are presented separately for the still images experiment (left columns), and the IVE test and retest experiments (middle and right columns, respectively). The "importance" columns show the relative importance of each factor (i.e., green, wide, and low traffic) as reflected by the coefficients sizes. The dimensions of aesthetics and enjoyment were similarly affected by the combinations of attributes that were examined. For both dimensions, greenery was by far the most dominant attribute in terms of its importance (70.6% for aesthetics and 64.3% for enjoyment). Cycle lane width and volume of pedestrian and cyclist traffic had a much more minor impact on sense of aesthetics and enjoyment (15.8% and 13.5% for aesthetics, respectively, and 28.0% and 7.7% for enjoyment, respectively), with coefficients mostly being statistically non-significant. This impact of greenery was even more substantial in the retest round (importance of 90.1% for aesthetics and 88.8% for enjoyment), with cycle lane width and traffic volume having a very minor impact on these dimensions of experience and with statistically non-significant coefficients.

As expected, green elements were much less significant in relation to sense of safety. The importance values calculated for the green attribute were 16.4% and 24.7% for the test and retest rounds, respectively, with a non-significant coefficient in the test round. In contrast, cycle lane width appeared to be a more dominant attribute in determining sense of safety (importance values of 56.6% and 49.2% for test and retest rounds). Traffic volume was also found to be more relevant to sense of safety than to the other dimensions of experience that were tested.

A comparison of the results of the test and retest conjoint experiments reveals some differences. As noted above, the green attribute had a more dominant influence on the cycling experience in the retest experiment. Table 3 includes the test–retest reliability scores of various common statistical indicators and tests (ICC (3,1), ICC (3,k), weighted kappa, Pearson's r, and hit rate). While ICC (3,1) scores serve as the basis for the comparison in this study (following Weir, 2005), the table includes the additional indicators to allow comparison with other studies in which different measures were used. An ICC score that denotes a good reliability level varies from one study to another, but is commonly greater than 0.6 and preferably above 0.8. In the current test–retest experiment, all ICC (3,1) scores are between 0.5 and 0.6, with aesthetics having the highest score (0.583) and enjoyment the lowest (0.504). Weighted kappas are all below 0.6 and hit rate is about one third in all dimensions of experience. Only the ICC (3,k) test presents a relatively good test–retest reliability close to or greater than 0.7. The rather low compliance between the two rounds indicates relatively poor repeatability.

4.2.2. Post-VR experiment ranking task

Table 4 presents the test–retest conjoint experiment results of the overall ranking. This conjoint questionnaire was administrated after the IVE experiment using still images with texts that described the attributes of the cycled environments (see, for example, Fig. 2). The results indicate that the presence of green elements had the strongest effect on determining overall attractiveness with importance values of 65.3% and 82.0% for the test and retest experiments, respectively. Cycle lane width was the second most important environmental attribute, with importance values of 19.9% for the test round and 13.1% for the retest. Traffic volume had the lowest effects and had an insignificant coefficient on the retest round. In this experiment, test–retest indicators were higher compared to the IVE experiments (ICC(3,1) = 0.681, ICC(3,k) = 0.810, wkappa = 0.600, hit rate = 61.5%), indicating an acceptable reliability level.

4.3. Differences between still images and IVE conjoint experiments

The left-hand columns of Table 3 display the results of the VR-equivalent still images experiment. A comparison between the environmental effects in this experiment and those in the IVE test round reveals a high level of similarity in the cases of aesthetics and enjoyment in terms of importance values. For example, the importance of greenery was 77.5% for the still images experiment, compared to 70.6% in the test experiment for sense of aesthetics. Similarly, the importance of greenery was 66.5% and 64.3% for the still images and test rounds, respectively, for the sense of enjoyment. In contrast, sense of safety showed dissimilar patterns. Traffic volume was of substantially lower importance (5.9%) in the still images experiment compared to the VR experiment (27%). This could be attributed to the dynamic representation of traffic in the VR, which generated a more realistic experience of insecurity that could not be achieved with still images and text alone.

Dissimilarities between the results of the still images and VR experiments were also observed in the ranking task. In the more cognitive still images experiment, cycling lane width was significantly more dominant (importance value of 31.9%) than the traffic volume attribute (1.1%) compared to the VR test experiment (19.9% for cycling lane and 14.9% for traffic). Again, this may indicate that the dynamic representation of traffic generates a stronger impact compared to static images.

A comparison between the results of the still images experiment and

Table 4

Conjoint experiment results of overall ranking of still images and post-IVE experiments (test and retest).

	Still images experiment				VR test experiment				VR retest experiment			
	Coefficients	t	p-value	Importance	Coefficients	t	p-value	Importance	Coefficients	t	p-value	Importance
(Constant)	1.283	13.703	< 0.001		1.535	13.959	< 0.001		1.675	15.193	< 0.001	
Green	1.653	17.435	< 0.001	67.1%	1.203	10.959	< 0.001	65.3%	1.340	12.176	< 0.001	82.0%
Wide	0.785	8.285	< 0.001	31.9%	0.367	3.339	0.001	19.9%	0.215	1.952	0.052	13.1%
Low traffic	0.026	0.276	0.783	1.1%	0.274	2.495	0.013	14.9%	0.080	0.723	0.470	4.9%
	R2(adjusted) = 0.619 (0.614)				R2(adjusted) = 0.293 (0.286)				R2(adjusted) = 0.370 (0.363)			
	N = 220 (55 respondents)				N = 336 (84 respondents)				N = 264 (66 respondents)			
					Test vs. retest comparison**ICC(3,1) = 0.681**, ICC(3,k) = 0.810, wkappa = 0.600, r = 0.680(Pvalue < 0.001), hit rate = 61.5%							

– Dependent variable = reverse ranking order whereby 1 was originally defined as the most attractive cycling environment and 4 the least attractive environment.

the VR retest experiment reveals only little similarity between the two setups. As our review above suggests, such differences between still and immersive displays were in fact expected.

5. Discussion

This article described the motivation behind and the physical setup of an SP experiment within IVE. While the results support the feasibility of employing this technique, some limitations and disadvantages should be acknowledged. First, the development of an SP experiment within IVE is currently costly and requires high levels of programming and graphic design skills, which are barriers to implementation. This is expected to improve in the coming years with the advancement of the technology and the introduction of more user-friendly VR tools and software. There are also inherent limitations concerning the research design and the data collection. In order to provide participants with an affective experience, they should be exposed to the studied environment for long periods. Though it is still not clear what a sufficient exposure is (and this will most likely vary from one study to another), it is expected that an exposure of at least several tens of seconds is required. This in turn increases the time it takes to administrate an IVE experiment compared to a more traditional SP task. Moreover, the negative side effects (or “VR sickness”)—such as nausea, headache, and disorientation—that some individuals experience after long exposures to VR display, limit how long researchers can expose non-experienced participants to IVEs to about 10 min. This might improve in the future as a result of technological improvements that will reduce the negative side effects, and due to the popularization of the technology, which will increase the number of VR-experienced participants, who can interact with IVEs for longer periods of time.

On the other hand, the added value of IVEs is in generating a sense of embodied experience, which is supposed to improve external validity (Patterson et al., 2017). The superior ITC–SOPI scores of the IVE experiment compared to the more traditional still images experiment support this idea. Future studies should employ a within-subject design in which the same participants are subject to the two experiments (i.e., still images and immersive). This might emphasize the differences between the immersive and dynamic environment and the static one-dimensional images. The combination of realistic environmental representations and the ability to fully control environmental characteristics is what makes this method so appealing. Because researchers would like to know that the results they obtain in the lab will be valid and usable in real life, there is a strong incentive to further develop the method, despite its inherent disadvantages. However, it should be noted that while the results of the study support the hypothesis that IVEs demonstrate higher level of realism, we can only assume, based on the literature (Slater & Wilbur, 1997) that this will in fact result in higher levels of external validity. In order to establish this relation more studies that compare SP task with real-life behavior need to be implemented. Another possible advantage of VR experiments is that, due to their novelty, they are likely to be more attractive to participants. This could make the recruitment of participants easier and the participants themselves more devoted to the task. However, the impact of the technology on compliance should be further examined in future studies.

An important observation concerning the sense of presence is the difference between experienced and non-experienced VR users. The former reported lower levels of presence, most likely because they were less overwhelmed by the technology and therefore more critical about imperfect representations of the virtual environments. While the excitement of non-experienced VR participants was not systematically recorded in the experiment, it was reflected in their enthusiastic behavior and in the comments they made to the research assistant during the experiments. However, while non-experienced participants reported higher levels of presence, it is most likely that their attention was highly distracted by their excitement, leading to the “contamination” of the results. The level of familiarity with VR should therefore be taken into account and monitored when performing IVE experiments.

The relatively low test–retest reliability scores that were obtained in the IVE rating experiments are surprising, given the robustness of the conjoint technique. Few hypotheses could be suggested to explain this. First, participants’ encounter with a novel cycling environment in the first experiment might have provoked different cognitive and affective responses compared to second experiment in which the environment was already familiar. This “novelty effect” might not be manifested when static, less realistic still images or text-only are used. Second, and related to the previous hypothesis, it is likely that during the first round, participants were preoccupied with mastering the technology and getting used to the VR experience, and thus paid substantially less attention to the experiment’s task itself. In the second round, participants were more familiar with the sensation of cycling within IVE. This in turn allowed them to be more engaged with the cycling experience and the rating task. The fact that the reliability indicators of the more traditional still images and text conjoint task yield acceptable repeatability, further supports the idea that the engagement with the virtual environment is the main cause for the low ICC scores in the rating task within the IVE. Third, while environments were distinguished from one another as reflected in the commonly significant factors, it could be that the environments do not generate a distinct cycling experience. Other factors (e.g. street cleanliness, and junctions) or attributes might be required to generate a more significant experience. The implementation of, for example, poorer cycling infrastructure (e.g., absence of cycling lanes) might have led to more distinct and reproducible impressions. Similarly, other types of green elements and volumes of pedestrians and cyclists might have had a more distinct impact on cycling experience.

The similarity between the still images scores and the first IVE experiment for the aesthetics and enjoyment is less clear, although similar results have been reported in the past (Orzechowski et al., 2005; Patterson et al., 2017). It could be that the effect of the first exposure to IVE in the test round prevented participants from having a full sense of embodied experience. Under these circumstances, participants might have engaged the virtual environment in a rather mechanical way, which led to the elicitation of cognitive assessments rather than more affective ones. This observation should be further examined in future studies. Lastly, it is important to note that still images experiments demonstrated better overall fit in terms of R^2 compared to the VR test experiment and in most cases also compared to VR retest experiments. It might be that the more complex environment that is displayed in IVEs generates noise which leads to more ambiguous results. Alternatively, and as discussed above, the first encounter with the cycling environment might have distracted participants’ attention and generated additional noise. It is therefore advisable to expose participants to the environments for longer periods of times or to allow them more training sessions when possible. This in turn might reduce the noise generated by the complex environment and first encounter with the new technology.

6. Conclusions

This study is one of the first attempts to systematically evaluate the implementation of an SP experiment within IVE (see also: Farooq et al., 2018). Although it was not always possible to extract conclusive results, several valuable insights and issues should be noted. First, IVE does seem to generate a greater sense of presence. This is reflected in the higher ITC–SOPI scores and in the more realistic results in the case of traffic volume. A greater level of realism of IVE compared to still images was also reported by Farooq et al. (2018). Second, familiarity with IVE seems to have a crucial impact on the results, as a first encounter with VR technology (controllers and display) may distract participants’ attention. As this study showed, short introductory tasks are not necessarily sufficient to overcome this effect. Nevertheless, it is believed that as the technology grows in popularity, this effect will decrease. Third, the results indicate that when a cognitive assessment of general environmental elements is required, traditional text and/or still images experiments might be a better solution due to their higher test–retest reliability, their lower costs, and their ease of implementation. The main advantage of IVE technology is in testing the more affective and dynamic dimensions of reactions to specific environmental scenarios and planned environmental alternatives, such as sense of security, crowdedness, and other sensations related to the flow of objects and people. Fourth, and related to the previous point, the ability to represent a dynamic environment is yet another important advantage of IVE technology. In the current study, dynamic aspects were reflected in the movement of cyclists, pedestrians, and the participants themselves; however, other dynamic elements—such as street life and sound and light features—could be considered in the future. Studies that are highly dependent on such elements should certainly consider taking advantage of IVE technology.

Based on the last two insights, we suggest the implementation of the following two-stage procedure in similar SP studies. The first stage should comprise a more traditional static SP experiment, so as to detect relevant prototypical environmental elements that are more cognitive in nature. This stage could be skipped in cases where the literature on the studied environmental phenomenon is well established and up to date. In the second stage, researchers can implement specific environmental attributes (e.g., particular tree species), preferably in a very specific environment (e.g., representation of an existing street), and if necessary integrate dynamic elements (Patterson et al., 2017).

The study also had several limitations that should be noted. First, the sample was non-representative: It did not include, for example, such important groups as older adults, adolescents, and the technologically illiterate. Second, while the test–retest experiment was conducted in a within-subject design, the comparison between still images and IVE could also benefit from a similar design. Third, only three environmental factors with two levels each were implemented in the present study. It is important to test the feasibility of integrating more complex environmental combinations.

VR technology allows researchers to implement more sophisticated tools that can support and, in the future, maybe even replace the traditional SP techniques. These may include walking and cycling speed and acceleration, the position of subjects relative to other objects such as people, buildings, and roads, head and eye movements, and the utilization of bio-sensors that can serve as indicators of emotional arousal. Moreover, future studies may integrate game elements such as choosing and searching tasks within the IVE, points earning, etc., which may motivate participants and generate more realistic reactions. Some gaming techniques may even be used to support the adoption of more healthy behaviors by participants (Baranowski et al., 2016). The utilization of immersive technology for research purposes is still in its infancy, and while these technologies hold great potential, researchers should always ensure that they understand the limitations and benefits that come with them.

Acknowledgments

The study was supported by the Netherlands Organisation for Scientific Research (NWO), grant number 485-14-005.

References

Alriksson, S., & Öberg, T. (2008). Conjoint analysis for environmental evaluation a review of methods and applications. *Environmental Science and Pollution Research, 15*(3), 244–257.

Anthes, C., Garcia-Hernandez, R. J., Wiedemann, M., & Kranzlmüller, D. (2016). State of the art of virtual reality technology. *2016 IEEE aerospace conference* (pp. 1–19). . https://doi.org/10.1109/AERO.2016.7500674.

Baranowski, T., Blumberg, F., Buday, R., DeSmet, A., Fiellin, L. E., Green, C. S., ... Young, K. (2016). Games for health for children—current status and needed research. *Games for Health Journal, 5*(1), 1–12. https://doi.org/10.1089/g4h.2015.0026.

Bateman, I. J., Day, B. H., Jones, A. P., & Jude, S. (2009). Reducing gain–loss asymmetry: A virtual reality choice experiment valuing land use change. *Journal of Environmental Economics and Management, 58*(1), 106–118. https://doi.org/10.1016/J.JEEM.2008.05.003.

Birenboim, A. (2018). The influence of urban environments on our subjective momentary experiences. *Environment and Planning B: Urban Analytics and City Science, 45*(5), 915–932. https://doi.org/10.1177/2399808317690149.

Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry, 13*(2), 103–124. https://doi.org/10.1207/S15327965PLI1302_01.

Borgers, A., Brouwer, M., Kunen, T., Jessurun Joran, J., & Janssen Ingrid, I. (2010). A virtual reality tool to measure shoppers' tenant mix preferences. *Computers, Environment and Urban Systems, 34*(5), 377–388. https://doi.org/10.1016/j.compenvurbsys.2010.04.002.

Bryan, S., Gold, L., Sheldon, R., & Buxton, M. (2000). Preference measurement using conjoint methods: An empirical investigation of reliability. *Health Economics, 9*(5), 385–395. https://doi.org/10.1002/1099-1050(200007)9:5<385::AID-HEC533>3.0.CO;2-W.

Burgoine, T., Forouhi, N. G., Griffin, S. J., Wareham, N. J., & Monsivais, P. (2014). Associations between exposure to takeaway food outlets, takeaway food consumption, and body weight in Cambridgeshire, UK: Population based, cross sectional study. *BMJ (Clinical Research Ed.), 348*, g1464. https://doi.org/10.1136/BMJ.G1464.

Caulfield, B., Brick, E., & McCarthy, O. T. (2012). Determining bicycle infrastructure preferences – A case study of Dublin. *Transportation Research Part D: Transport and Environment, 17*(5), 413–417. https://doi.org/10.1016/J.TRD.2012.04.001.

Cervero, R. (2002). Built environments and mode choice: Toward a normative framework. *Transportation Research Part D: Transport and Environment, 7*(4), 265–284. https://doi.org/10.1016/S1361-9209(01)00024-4.

Cosma, G., Ronchi, E., & Nilsson, D. (2016). Way-finding lighting systems for rail tunnel evacuation: A virtual reality experiment with Oculus Rift®. *Journal of Transportation Safety & Security, 8*(1), 101–117. https://doi.org/10.1080/19439962.2015.1046621.

de Kort, Y. A. W., Meijnders, A. L., Sponselee, A. A. G., & IJsselsteijn, W. A. (2006). What's wrong with virtual trees? Restoring from stress in a mediated environment. *Journal of Environmental Psychology, 26*(4), 309–320. https://doi.org/10.1016/j.jenvp.2006.09.001.

Dijkstra, J., van Leeuwen, J., & Timmermans, H. (2003). Evaluating design alternatives using conjoint experiments in virtual reality. *Environment and Planning B: Planning and Design, 30*(3), 357–367. https://doi.org/10.1068/b12932.

Earnhart, D. (2002). Combining revealed and stated data to examine housing decisions using discrete choice analysis. *Journal of Urban Economics, 51*(1), 143–169. https://doi.org/10.1006/JUEC.2001.2241.

Evans, G. W. (2003). The built environment and mental health. *Journal of Urban Health: Bulletin of the New York Academy of Medicine, 80*(4), 536–555. https://doi.org/10.1093/jurban/jtg063.

Farooq, B., Cherchi, E., & Sobhani, A. (2018). Virtual immersive reality for stated preference travel behavior experiments: A case study of autonomous vehicles on urban roads. *Transportation Research Record: Journal of the Transportation Research Board, 2672*(50), 35–45. https://doi.org/10.1177/0361198118776810.

Fox, J., Arena, D., & Bailenson, J. N. (2009). Virtual reality: A survival guide for the social scientist. *Journal of Media Psychology, 21*(3), 95–113. https://doi.org/10.1027/1864-1105.21.3.95.

Fraser, S. D. S., & Lock, K. (2011). Cycling for transport and public health: A systematic review of the effect of the environment on cycling. *European Journal of Public Health, 21*(6), 738–743. https://doi.org/10.1093/eurpub/ckq145.

Gaggioli, A., Bassi, M., & Delle Fave, A. (2003). Quality of experience in virtual environments. In R. Giuseppe, D. Fabrizio, & I. A. Wijnand (Eds.). *Being there: Concepts, effects and measurements of user presence in synthetic environments* (pp. 121–135). Amsterdam: IOS press.

Ghekiere, A., Deforche, B., Mertens, L., De Bourdeaudhuij, I., Clarys, P., de Geus, B., ... Page, M. (2015). Creating cycling-friendly environments for children: Which micro-scale factors are most important? An experimental study using manipulated photographs. *PLoS ONE, 10*(12), e0143302. https://doi.org/10.1371/journal.pone.0143302.

Gorini, A., Capideville, C. S., De Leo, G., Mantovani, F., & Riva, G. (2011). The role of immersion and narrative in mediated presence: The virtual hospital experience. *Cyberpsychology, Behavior, and Social Networking, 14*(3), 99–105. https://doi.org/10.1089/cyber.2010.0100.

Grechkin, T. Y., Chihak, B. J., Cremer, J. F., Kearney, J. K., & Plumert, J. M. (2013). Perceiving and acting on complex affordances: How children and adults bicycle across two lanes of opposing traffic. *Journal of Experimental Psychology: Human Perception and Performance, 39*(1), 23–36. https://doi.org/10.1037/a0029716.

Handy, S., van Wee, B., & Kroesen, M. (2014). Promoting cycling for transport: Research needs and challenges. *Transport Reviews, 34*(1), 4–24. https://doi.org/10.1080/01441647.2013.860204.

Jansen, S., Boumeester, H., Coolen, H., Goetgeluk, R., & Molin, E. (2009). The impact of including images in a conjoint measurement task: Evidence from two small-scale studies. *Journal of Housing and the Built Environment, 24*(3), 271–297. https://doi.org/10.1007/s10901-009-9149-x.

Jansen, M., Kamphuis, C. B. M., Pierik, F. H., Ettema, D. F., & Dijst, M. J. (2018). Neighborhood-based PA and its environmental correlates: A GIS- and GPS based cross-sectional study in the Netherlands. *BMC Public Health, 18*(1), 233. https://doi.org/10.1186/s12889-018-5086-5.

Kahlert, D., & Schlicht, W. (2015). Older people's perceptions of pedestrian friendliness and traffic safety: An experiment using computer-simulated walking environments.

International Journal of Environmental Research and Public Health, 12(8), 10066–10078. https://doi.org/10.3390/ijerph120810066.

Katoshevski, R., & Timmermans, H. (2001). Using conjoint analysis to formulate user-centred guidelines for urban design: The example of new residential development in Israel. *Journal of Urban Design, 6*(1), 37–53. https://doi.org/10.1080/13574800120032888.

Kaźmierczak, A. (2013). The contribution of local parks to neighbourhood social ties. *Landscape and Urban Planning, 109*(1), 31–44. https://doi.org/10.1016/j.landurbplan.2012.05.007.

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.

Krizek, K. J., Handy, S. L., & Forsyth, A. (2009). Explaining changes in walking and bicycling behavior: Challenges for transportation research. *Environment and Planning B: Planning and Design, 36*(4), 725–740. https://doi.org/10.1068/b34023.

Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J. (2001). A cross-media presence questionnaire: The ITC-sense of presence inventory. *Presence, 10*(3), 282–297. https://doi.org/10.1162/105474601300343612.

Loomis, J. M., Blascovich, J. J., & Beall, A. C. (1999). Immersive virtual environment technology as a basic research tool in psychology. *Behavior Research Methods, Instruments, & Computers, 31*(4), 557–564. https://doi.org/10.3758/BF03200735.

Louvière, J. (1979). Modeling individual residential preferences: A totally disaggregate approach. *Transportation Research Part A: General, 13*(6), 373–384. https://doi.org/10.1016/0191-2607(79)90001-3.

Louvière, J., Hensher, D. A., & Swait, J. D. (2000). *Stated choice methods: Analysis and applications.* Cambridge University Press.

Louvière, J., & Timmermans, H. (1990). Stated preference and choice models applied to recreation research: A review. *Leisure Sciences, 12*(1), 9–32. https://doi.org/10.1080/01490409009513087.

Molin, E. J. E., Oppewal, H., & Timmermans, H. J. P. (2001). Analyzing heterogeneity in conjoint estimates of residential preferences. *Journal of Housing and the Built Environment, 16*(3–4), 267–284. https://doi.org/10.1023/A:1012539415696.

Morrongiello, B. A., Corbett, M., Milanovic, M., Pyne, S., & Vierich, R. (2015). Innovations in using virtual reality to study how children cross streets in traffic: Evidence for evasive action skills. *Injury Prevention, 21*(4), 266–270. https://doi.org/10.1136/injuryprev-2014-041357.

Nash, E. B., Edwards, G. W., Thompson, J. A., & Barfield, W. (2000). A review of presence and performance in virtual environments. *International Journal of Human-Computer Interaction, 12*(1), 1–41. https://doi.org/10.1207/S15327590IJHC1201_1.

Nazemi, M., van Eggermond, M., Erath, A., Schaffner, D., & Axhausen, K. W. (2018). *Perception of Speed and Space in Virtual Environments. Presented in: The 7th International Cycling Safety Conference (ICSC 2018). Barcelona*.

O'Hern, S., Oxley, J., & Stevenson, M. (2017). Validation of a bicycle simulator for road safety research. *Accident Analysis & Prevention, 100*, 53–58. https://doi.org/10.1016/J.AAP.2017.01.002.

Orme, B. K. (2010). *Getting started with conjoint analysis: Strategies for product design and pricing research.* Research Publishers.

Orzechowski, M. A., Arentze, T. A., Borgers, A. W. J., & Timmermans, H. J. P. (2005). Alternate methods of conjoint analysis for estimating housing preference functions: Effects of presentation style. *Journal of Housing and the Built Environment, 20*(4), 349–362. https://doi.org/10.1007/s10901-005-9019-0.

Patterson, Z., Darbani, J. M., Rezaei, A., Zacharias, J., & Yazdizadeh, A. (2017). Comparing text-only and virtual reality discrete choice experiments of neighbourhood choice. *Landscape and Urban Planning, 157*, 63–74. https://doi.org/10.1016/j.landurbplan.2016.05.024.

Poelman, M., Strak, M., Schmitz, O., Hoek, G., Karssenberg, D., Helbich, M., ... Vaartjes, I. (2018). Relations between the residential fast-food environment and the individual risk of cardiovascular diseases in The Netherlands: A nationwide follow-up study. *European Journal of Preventive Cardiology, 25*(13), 1397–1405. https://doi.org/10.1177/2047487318769458.

Rao, V. R. (2014). *Applied conjoint analysis*Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-87753-0.

Risto, M., & Martens, M. H. (2014). Driver headway choice: A comparison between driving simulator and real-road driving. *Transportation Research Part F: Traffic Psychology and Behaviour, 25*, 1–9. https://doi.org/10.1016/J.TRF.2014.05.001.

Ronchi, E., Kinateder, M., Müller, M., Jost, M., Nehfischer, M., Pauli, P., & Mühlberger, A. (2015). Evacuation travel paths in virtual reality experiments for tunnel safety analysis. *Fire Safety Journal, 71*, 257–267. https://doi.org/10.1016/j.firesaf.2014.11.005.

Rumschlag, G., Palumbo, T., Martin, A., Head, D., George, R., & Commissaris, R. L. (2015). The effects of texting on driving performance in a driving simulator: The influence of driver age. *Accident Analysis & Prevention, 74*, 145–149. https://doi.org/10.1016/J.AAP.2014.10.009.

Ruotolo, F., Maffei, L., Di Gabriele, M., Iachini, T., Masullo, M., Ruggiero, G., & Senese, V. P. (2013). Immersive virtual reality and environmental noise assessment: An innovative audio–visual approach. *Environmental Impact Assessment Review, 41*, 10–20. https://doi.org/10.1016/J.EIAR.2013.01.007.

Saelens, B. E., Sallis, J. F., & Frank, L. D. (2003). Environmental correlates of walking and cycling: Findings from the transportation, urban design, and planning literatures. *Annals of Behavioral Medicine*. https://doi.org/10.1207/S15324796ABM2502_03.

Sanchez-Vives, M. V., & Slater, M. (2005). From presence to consciousness through virtual reality. *Nature Reviews Neuroscience, 6*(4), 332–339. https://doi.org/10.1038/nrn1651.

Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence: Teleoperators & Virtual Environments, 6*(6), 603.

Smith, J. W. (2015). Immersive virtual environment technology to supplement environmental perception, preference and behavior research: A review with applications. *International Journal of Environmental Research and Public Health, 12*(9), 11486–11505. https://doi.org/10.3390/ijerph120911486.

Sylcott, B., Orsborn, S., & Cagan, J. (2016). The effect of product representation in visual conjoint analysis. *Journal of Mechanical Design, 138*(10), 101104. https://doi.org/10.1115/1.4034085.

Tilahun, N. Y., Levinson, D. M., & Krizek, K. J. (2007). Trails, lanes, or traffic: Valuing bicycle facilities with an adaptive stated preference survey. *Transportation Research Part A: Policy and Practice, 41*(4), 287–301. https://doi.org/10.1016/j.tra.2006.09.007.

Van Cauwenberg, J., Clarys, P., De Bourdeaudhuij, I., Ghekiere, A., de Geus, B., Owen, N., & Deforche, B. (2018). Environmental influences on older adults' transportation cycling experiences: A study using bike-along interviews. *Landscape and Urban Planning, 169*, 37–46. https://doi.org/10.1016/j.landurbplan.2017.08.003.

Vedel, S. E., Jacobsen, J. B., & Skov-Petersen, H. (2017). Bicyclists' preferences for route characteristics and crowding in Copenhagen – a choice experiment study of commuters. *Transportation Research Part A: Policy and Practice, 100*, 53–64. https://doi.org/10.1016/j.tra.2017.04.006.

Verhoeven, H., Ghekiere, A., Van Cauwenberg, J., Van Dyck, D., De Bourdeaudhuij, I., Clarys, P., & Deforche, B. (2017). Which physical and social environmental factors are most important for adolescents' cycling for transport? An experimental study using manipulated photographs. *International Journal of Behavioral Nutrition and Physical Activity, 14*(1), 108. https://doi.org/10.1186/s12966-017-0566-z.

Waller, D., Beall, A. C., & Loomis, J. M. (2004). Using virtual environments to assess directional knowledge. *Journal of Environmental Psychology, 24*(1), 105–116. https://doi.org/10.1016/S0272-4944(03)00051-3.

Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *Journal of Strength and Conditioning Research, 19*(1), 231–240.

Wendel-Vos, G. C. W., Schuit, A. J., De Niet, R., Boshuizen, H. C., Saris, W. H. M., & Kromhout, D. (2004). Factors of the physical environment associated with walking and bicycling. *Medicine and Science in Sports and Exercise, 36*(4), 725–730. https://doi.org/10.1249/01.MSS.0000121955.03461.0A.

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments, 7*(3), 225–240. https://doi.org/10.1162/105474698565686.

Yamada, I., & Thill, J.-C. (2003). Enhancing stated preference surveys in transportation research: The contribution of geographic information systems. *Transportation Planning and Technology, 26*(5), 377–396. https://doi.org/10.1080/03081060310001635850.

Zhao, Y., Dijst, M., & Chai, Y. (2016). Between haven and heaven in cities: A comparison between Beijing (China) and Utrecht (the Netherlands). *Urban Studies, 53*(12), 2469–2487. https://doi.org/10.1177/0042098015596106.