A virtual reality experiment to study citizen perception of future street scenarios

Javier Argota Sánchez-Vaquerizo[1,*], Carina I. Hausladen[1,2], Sachit Mahajan[1], Marc Matter[1], Michael Siebenmann[1], Michael A. B. van Eggermond[3], and Dirk Helbing[1,4]

[1]ETH Zürich, Computational Social Science, Zürich, 8092, Switzerland
[2]California Institute of Technology, Behavioral Economics, Pasadena, CA 91125, United States
[3]University of Applied Sciences Northwestern Switzerland (FHNW), Muttenz, 4132, Switzerland
[4]Complexity Science Hub, Vienna, 1080, Austria
[*]javier.argota@gess.ethz.ch

ABSTRACT

The current allocation of street space is based on expected vehicular peak-hour flows. Flexible and adaptive use of this space can respond to changing needs. To evaluate the acceptance of flexible street layouts, several urban environments were designed and implemented in virtual reality. Participants explored these designs in immersive virtual reality in a 2x3 mixed factorial experiment, in which we analysed self-reported, behavioural and physiological responses from participants. Distinct communication strategies were varied between subjects. Participants' responses reveal a preference for familiar solutions. Unconventional street layouts are less preferred, perceived as unsafe and cause a measurably greater stress response. Furthermore, information provision focusing on comparisons led participants to focus primarily on the drawbacks, instead of the advantages, of novel scenarios. When being able to freely express thoughts and opinions, participants were focused more on the impact of the space on behaviour rather than the objective physical features themselves. Especially this last finding suggests that it is vital to develop new street scenarios in an inclusive and democratic way: the success of innovating urban spaces depends on how well the vast diversity of citizens' needs is considered and met.

Introduction

Rapid urbanization has resulted in the majority of the world's population living in cities. At the same time, car-centred planning has resulted in the fact that up to one-third of the urban area is allocated to infrastructures for motor vehicles, including roads and parking lots[1]. With increasing population and employment densities within a limited area, cities face numerous challenges, ranging from air and noise pollution over congestion to affordable housing prices and suitable allocation of space[2].

The trend to provide more of this space to sustainable modes of transport and active mobility has recently resulted in tensions around the allocation and use of road space[3]. At the same time, the growth of cities comes with a growing desire for *public* space. Increasingly, streets are not only viewed as thoroughfares but as places. An allocation of roads to transport only conflicts with this desire[4]. However, there may be new solutions. The current allocation of space to motorized transport is based on expected peak-hour flows, while the demand for different modes of transport and uses largely varies with the time-of-day and day-of-week[3]. Therefore, underutilized spaces could be used for other purposes[5]. For example, urban planning concepts such as shared spaces or shared streets integrate vehicular traffic into social spaces by removing typical design elements such as signs and speed humps[6]. As we will see, novel technologies enable even better solutions.

Recently, within the framework of responsive, or adaptable, infrastructures, one explores flexible usage patterns of roads. With the emergence of digitally upgraded infrastructures and autonomous vehicles, future layouts and scenarios become possible in the near future, such as dynamic reversible lanes, laneless roads, or curbless flat streets[7–10]. Therefore, our study aims at investigating new street designs that enable adaptive usage paradigms for urban spaces, envisioning a future where there is no static distinction between roadways, sidewalks, and parking lots[11], but where streets can be used dynamically and flexibly[12]. We would like to emphasize that the term "street" pertains to the entire space located between two buildings and not just limited to the roadways.

So far, however, there is limited knowledge on how flexible streetscapes will be designed, operated, and accepted by people[13]. Among others, it will be essential to study movement patterns of pedestrians in relation to new environments[14]. An understanding of how street design in urban settings affects the behaviour of people can inform the creation of safe urban environments and can support the decision-making of stakeholders when new mobility paradigms arise, including policy-makers, experts, and citizens.

Classical segregated lanes may maximize speeds for motorized vehicles but do not necessarily enhance the overall multimodal level-of-service. Future streets and vehicle technologies have the potential to overcome this segregation and promote a higher diversity of uses on streets depending on changing needs[13,15]. However, a radical shift in street design may not be accepted by drivers, as they are used to being prioritized. Already, changes to the allocation of space (e.g. reduction of lanes or parking lots, traffic calming, pedestrianization, expansion of transit and bikes lanes) or in pricing schemes (e.g. parking fees, mobility pricing) often face resistance[16]. Therefore, determining the acceptability of measures to be taken and their suitable communication are crucial for successful changes in street design or operation.

For this, it is important to consider the perspectives of all road users and provide them with incentives to embrace the changes in street design and operation, such as a reduction in speed in favour of interoperability and reduced accident severity. *Our study proposes that intrinsic motivation to cooperate can be enhanced by perceived improvements in the quality of mobility.* Implementing changes in traffic flows that promote a slower yet more continuous traffic movement may help to improve the acceptability of speed limits for car drivers. Moreover, such changes may enhance the safety of all road users, promote shared use of the street, and create a more equitable and sustainable transportation system. By prioritizing the needs of all road users and emphasizing the importance of accessibility, comfort, and safety, one may enhance the overall livability of urban areas and foster a more inclusive and sustainable community.

Current engineering efforts to improve autonomous driving focus, for example, on enhancing sensing capabilities, intervehicle and car-to-infrastructure communication, vehicle safety, and trust in navigation[17–19]. However, the impact of disruptive mobility paradigms, on urbanism and traffic, such as adaptive patterns of street usage, is not yet well understood. To accurately assess the impact of disruptive mobility technologies, it is necessary to gather feedback from human subjects, particularly pedestrians. Specifically, traffic planners will need to understand how *pedestrians* perceive these technologies and paradigms, and whether they are seen as a safety threat, causing psychological distress, or even promoting a variable and diverse use of urban spaces. Therefore, this study focuses on understanding the perception of these technologies by pedestrians and their impact on urban spaces. We aim to provide new insights that can inform future urban planning and design decisions and ensure that the needs and perspectives of all road users are taken into account.

Virtual reality (VR), as a research tool, has been used previously to assess crossing behaviour[20], study interaction between pedestrians and autonomous vehicles[21], communicate future urban designs[22] and assess the willingness-to-cycle[21]. The advantages of VR are, among others, that it is possible to communicate future designs and technology in a realistic, immersive and safe fashion, against relatively low costs. Furthermore, comparison between real and virtual environments reveals that individuals make similar decisions when studying crossing behaviour[23], pedestrian behaviour matches real-life norms[24] and VR can aid in imagining future technology when compared to text-based surveys[25].

Our experiment has three primary goals. First, we aim to test new street scenarios that address the challenges discussed above and promote a more sustainable approach to transportation. Second, we aim to investigate how these scenarios are perceived by the public, including their perceived safety, and how they impact travel patterns. Third, we aim to test how novelties in street design can be effectively communicated to the public. Our focus is on examining the perceived safety of various street designs and understanding the psychological responses to these settings, as they significantly influence travel patterns. By doing so, we aim to facilitate a shift in the prevailing mobility culture, which currently gravitates around motorized vehicles, towards a more eco-friendly and sustainable approach to street usage. Through our research, we hope to provide insights that can inform future urban planning and design decisions and promote the adoption of more inclusive and sustainable transportation policies.

To achieve our goals, we design and conduct a VR experiment, which focuses on the pedestrian perspective (Figure 1). Specifically, we test three different street design scenarios that emphasize the concept of shared space and incorporate emerging technologies such as smart pavements and autonomous vehicles, which will become usable in the medium-term perspective. This is our first goal. In our VR experiments, participants are asked to cross a virtual street in three different virtual environments (within-subject factor), each representing a different street design scenario. Through this, we aim to quantify user perception and understand the impact of these scenarios on safety and travel patterns. To achieve this, we collect three types of data: self-reported preferences, behavioural data and physiological responses, including heart rate as well as head and body movement patterns. This is our second goal. To investigate how these changes should be best communicated to the public, we perform a between-subject treatment. Half of the participants watch the new traffic scenarios sequentially in an isolated way, while the other half watch the new street paradigms in parallel on a split screen. We hypothesize that the split-screen approach supports comparison and helps to recognize the advantages and disadvantages of each scenario better than watching videos sequentially. Ultimately, the insights we gain from this study can inform the design and implementation of future transportation policies that are more inclusive and sustainable. This is our third goal.

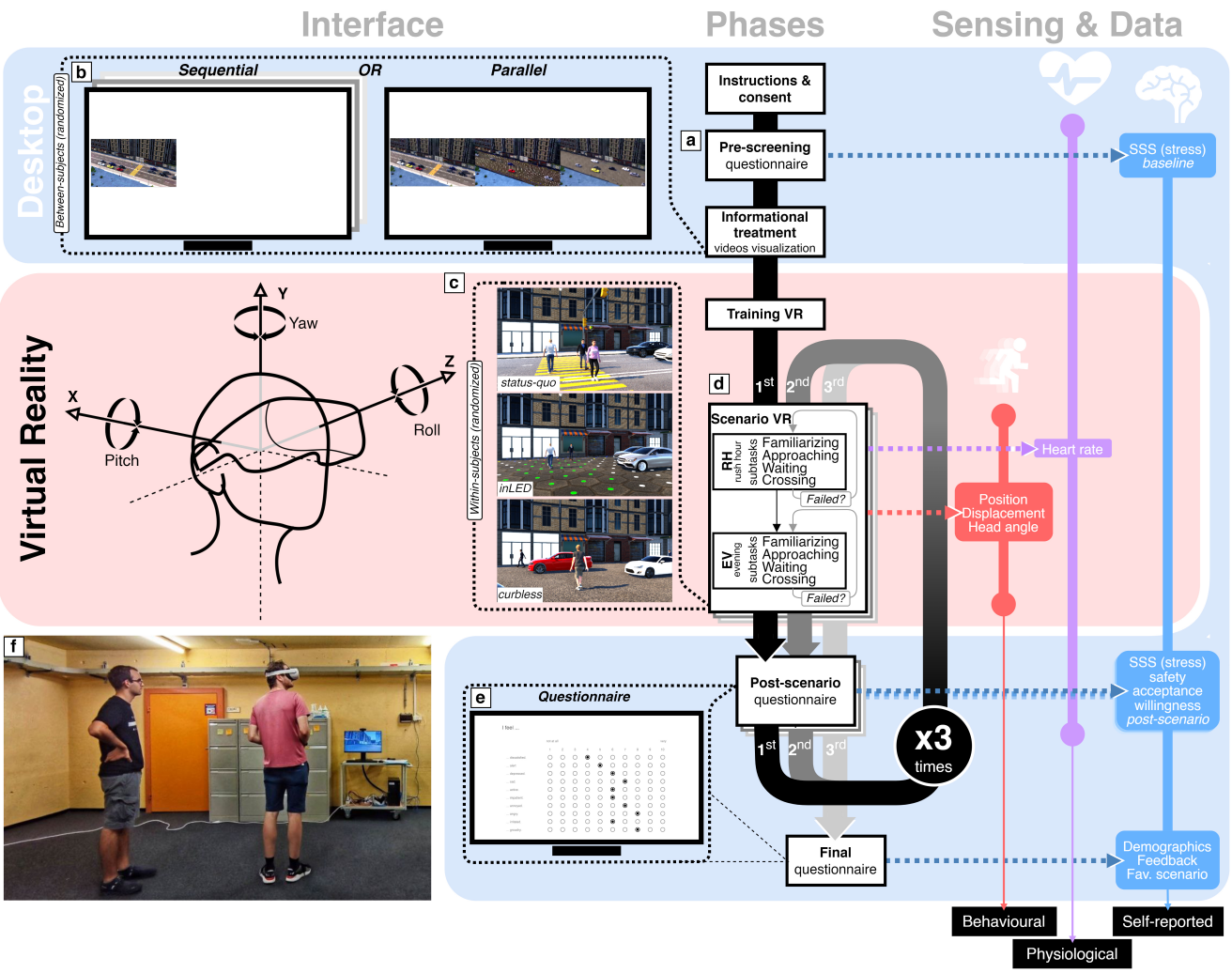

Figure 1. Experiment design, phases, devices and data collection during the experiment setting.

Results

The study involved 43 participants with a mean age of 24.76 (±3.73) years, comprising 62.8% (n=27) female individuals. The majority of the participants were students.

Table 1. Summary of participants and recorded sessions. Rows display the number of participants per information treatment group (*sequential*, *parallel*). Columns in the right-most part of the table show the number of sessions per scenario. Each participant experienced the three scenarios (*status-quo*, *inLED*, *curbless*), in two different configurations (morning rush hour, and evening). Participants were given at most two attempts to cross each configuration in each scenario, which resulted in a different total number of sessions per scenario and per treatment, depending on the number of failed attempts.

		individual participants	**sessions** scenarios			
			status-quo	*inLED*	*curbless*	*TOTAL*
information treatments	*sequential*	21	42	47	42	131
	parallel	22	44	48	43	135
	TOTAL	**43**	86	95	85	**266**

Number of crossing attempts Participants were briefed to cross the road safely (i.e. without being hit by a car), with each participant given two attempts per configuration in each scenario to do so (**d** in Figure 1). Table 1 displays the number of participants by scenario (columns) and treatment (rows). Overall, with only 4.5% there was a low rate of failed attempts, but there was a significant impact of the scenario type on the failure rate (n=266, $p_{two-sided\ Fisher} = 1.1e-04$). Most of the failures occurred in scenario *inLED*. Apart from this, there were no significant differences in failure rates observed across treatments (*sequential* or *parallel*), indicating that this intervention had no significant impact on performing the task (n=266, $p_{two-sided\ Fisher} = 1.0$).

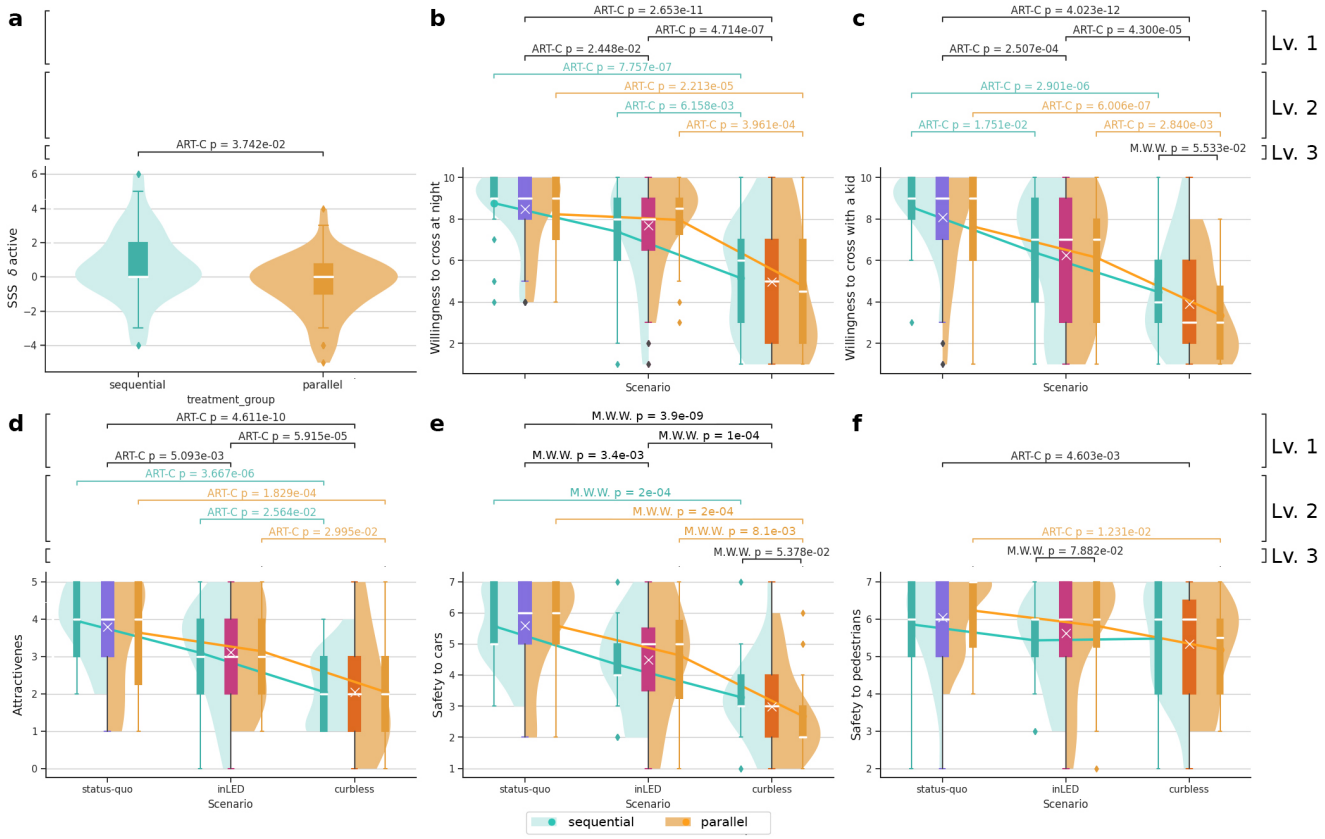

Figure 2. Self-reported responses in the structured survey. Each panel presents a different response to the structured survey, which showed statistically significant differences ($p < 0.05$) between *information treatments*, *scenarios*, or their interaction (Information treatments × Scenario). Panel **a** displays the distributions for the self-reported "active" feeling on the Short Stress State (SSS) questionnaire. Panels **b-f** show responses with statistically significant differences for *scenarios* and interactions between interventions. When shown, scenarios are grouped on the X-axis, featuring the aggregated distribution as a coloured (blue, magenta, bright orange) box plot. Mean values are denoted by a white ×. Violin plots show distributions for each information treatment within the same scenario: mint for *sequential* and light orange for *parallel*. They are complemented by corresponding box plots displaying median values and interquartile ranges. Lines linking information treatments boxplots show mean values visualizing within-treatment changes between scenarios. Annotations[26] on top of each panel display the contrast multifactor pairwise posthoc tests for the three levels of analysis after performing an Aligned Rank Transform (ART) ANOVA test[27] (or alternatively for robustness a Kruskal-Wallis H test[28]). Only significant results ($p < 0.05$) resulting from using the Aligned Rank Transform Contrast (ART-C) algorithm[29] (or alternatively for robustness Mann-Whitney-Wilcoxon) are reported, adjusted for the multiple comparisons with the Benjamini-Yekutieli method[30]. Additionally, for the within-scenario pairwise contrast tests (level 3), a one-sided Mann-Whitney-Wilcoxon test is used, with a significance level $p < 0.1$, for its higher power due to the particular characteristics of the compared distributions[31].

Stress, safety, willingness to cross, and attractiveness Directly after completing the crossing for each scenario, participants responded to a brief structured survey. This survey includes a Short Stress State (SSS)[32] questionnaire, and questions regarding safety, willingness and attractiveness perception. To address repeated measures of non-parametric variables, we employed an Aligned Rank Transform (ART) ANOVA test[27] (see Materials and Methods for more details). Our analysis is structured across

three levels: design scenario (Level 1), information treatment × scenario (Level 2), and information treatment (Level 3).

Scenarios (Level 1) elicit significant variations in responses. Participants' reactions to the scenarios follow a ranking pattern of *status-quo*>*inLED*>*curbless*, suggesting that participants are less willing to cross unconventional street layouts, these are perceived as less attractive and as less safe. This trend is visually apparent through the negative slopes of the lines in panels **b**, **c**, **d**, and **e**. In panels **e** and **f**, which analyze the differences in perceived safety for pedestrians and cars, differences are observed. Panel **e** shows a steep slope with significant scenario comparisons (*status-quo*>*inLED*>*curbless*), while panel **f** exhibits nearly horizontal lines with only one significant comparison (*status-quo*=*inLED*>*curbless*). This suggests that the aversion to unconventional scenarios is caused mainly due to the proximity of cars rather than the proximity to pedestrians.

Analyzing the scenarios within information treatments (Level 2) is facilitated by counting the mint and light orange brackets depicted on annotations on top of panels of Figure 2. We observe that scenario *status-quo* receives significantly higher ratings than scenario *curbless* in more questions posed to the *sequential* group (4 questions) compared to the *parallel* group (5 questions). Similarly when comparing scenarios *inLED* and *curbless*: the *sequential* group perceives scenario *curbless* as less likely to cross the street at night and less attractive than scenario *inLED*. In addition to these observations, the *parallel* group is additionally less inclined to cross the street with a child and believes the street is less safe regarding interaction with cars.

The information treatment (level 3) demonstrated an impact on three additional questions: In scenario *curbless*, participants belonging to the *parallel* group exhibited a reduced willingness to cross the street with a child and perceived the street as less safe regarding the interaction with cars compared to those in the *sequential* group. In essence, the *parallel* group perceived scenario *curbless* more negatively than the *sequential* group. Conversely, in scenario *inLED*, participants in the *parallel* group considered the street to be safer for pedestrians than those in the *sequential* group. In summary, Level 3 differences are most frequently observed in scenario *curbless*, where factors such as willingness to cross the street with a child and perceived car safety are influenced by the presentation of information regarding the scenarios. We delve deeper into this topic in the discussion section.

Duration The largest differences concerning the *duration* of the experiment are found overall in the subtasks approaching and crossing (panels **b**, **c** and **e** in Figure 3). The results indicate that participants in the *parallel* treatment took significantly longer to approach the crosswalk compared to those in the *sequential* treatment ($N = 266$, $p_{adj\ BY} = 3.583e-03$). Moreover, this finding remained consistent when examining the data within the *inLED* scenario ($p_{adj\ BY} = 2.271e-02$) instead of the aggregate level. Similarly, the duration of waiting (panel **d** in Figure 3) was highest in the *inLED* scenario ($p_{adj\ BY} = 7.696e-03$). The general trend of increased duration for the *inLED* scenario was also reflected in the duration of the whole experiment (panel **f** in Figure 3, $p_{adj\ BY} = 3.448e-03$). The findings for crossing time (panel **e**) were in contrast to those for waiting and approaching. Specifically, participants took significantly longer to cross in the *curbless* scenario compared to both the *status-quo* ($p_{adj\ BY} = 2.116e-07$) and *inLED* ($p_{adj\ BY} = 4.679e-06$) scenarios. This result is consistent with the fact that the *curbless* scenario was the only one without a traffic light or in-pavement LEDs, allowing participants to take as much time as needed to cross. This could potentially explain the longer crossing times observed in panel **e**.

In comparing treatments across scenarios (level 2), the results indicate that the *sequential* treatment exhibited a strict ordering of *status-quo* < *inLED* < *curbless*. Likewise, the *parallel* treatment showed an ordering of *status-quo* ≈ *inLED* < *curbless*.

Displacement In the following, we examine the participants' *movement patterns*, which encompassed both their actual physical movements in a 5m × 12m sized room and their ability to "teleport" within the VR environment. We first calculated the participants' *speed*, which was defined as meters moved per second. The standard deviation of speed (σ_{speed}) served as a proxy for the uniformity in movement pace. Panel **g** in Figure 3 shows that σ_{speed} differed significantly across treatments in the *curbless* scenario. Specifically, participants in the *sequential* treatment exhibited more variation in their movement speed compared to those in the *parallel* treatment. Our analysis of the overall distance walked (panel **h** in Figure 3) revealed the following ordering: *status-quo* < *inLED* ≈ *curbless*. This effect was primarily driven by the *parallel* treatment, as we observed a significant difference in distance walked between the *status-quo* and *inLED* scenarios for this treatment.

Head movement During the VR experiment, *head movement and displacement* was logged to quantify the exploration of the environment[33–35]. The head movement for scenario *curbless* was not recorded due to a bug. Consequently, we only report significant differences between scenarios *status-quo* and *inLED*. Our analysis shows that pitch (i.e., looking up and down) and yaw (i.e., turning head left and right) differed significantly across the treatments in terms of both, their mean (μ) and standard deviation (σ) (panels **i-l** and **e** in Figure 3). Specifically, participants looked less up and down in the *inLED* scenario compared to the *status-quo* scenario, while they turned their head more from left to right in the *inLED* scenario compared to the *status-quo* scenario. In other words, the scenario appeared to influence participants' visual exploration patterns: the *inLED* scenario required participants to focus more on exploring the ground plane, while the *status-quo* scenario needed more exploration between the ground and eye-level plane.

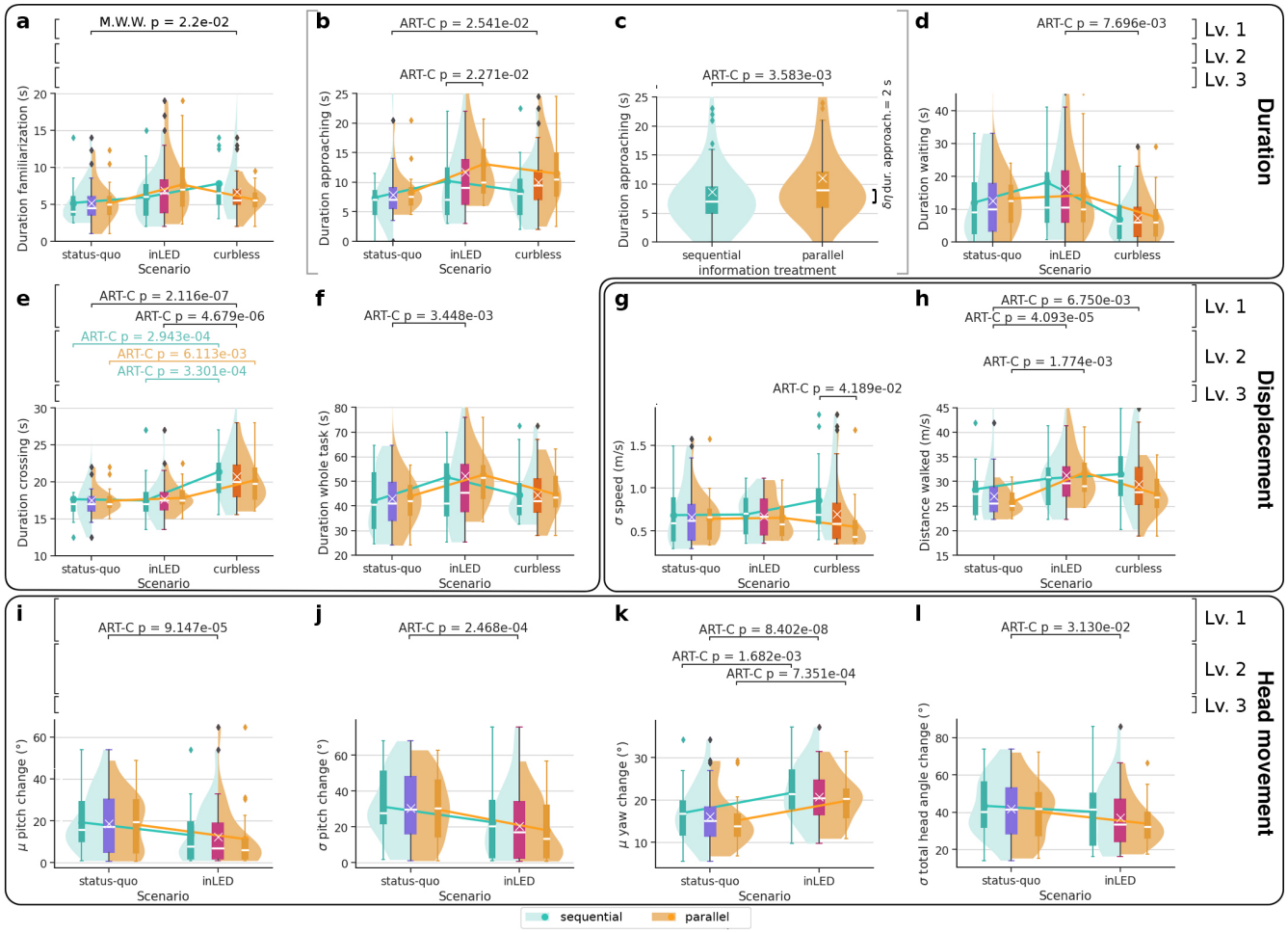

Figure 3. Behavioural responses. Each panel shows a different response measured from the participants' behaviour which showed statistically significant differences ($p < 0.05$) between *information treatments*, *scenarios*, or their interaction (Information treatments × Scenario). The participant's activity was divided into four subtasks, namely *familiarization* with the virtual environment (**a**), *approaching* the crossing (**b**-**c**), *waiting* to cross (**d**), and *crossing* (**e**). Duration in seconds is shown in the y-axis for each subtask in panels **a**-**e**, and for the whole task in panel **f**. Panels **g** and **h** illustrate differences in uniformity of speed (measured as σ_{speed}) and total walked distance within the VR environment. Panels **i**-**l** highlight differences in head movement. Specifically, we report the mean (μ) and standard deviation (σ) of change in pitch (i.e., looking up or down, see panels **i** and **j**) and yaw (i.e., turning head left or right, see panels **k**) as well as the cumulative total angle change in direction of the three axes per second (panel **l**). When shown, scenarios are grouped on the X-axis, featuring the aggregated distribution as a coloured (blue, magenta, bright orange) box plot. Mean values are denoted with a white ×. Violin plots show distributions for each information treatment within the same scenario: mint for *sequential* and orange for *parallel*. They are complemented by corresponding box plots displaying median η values and interquartile ranges. Lines linking information treatments boxplots show mean values visualizing within-treatment changes between scenarios. Annotations[26] on top of each panel display the contrast multifactor pairwise posthoc tests for the three levels of analysis after performing an Aligned Rank Transform (ART) ANOVA test[27] (or alternatively for robustness a Kruskal-Wallis H test[28]). Only significant results ($p < 0.05$) resulting from using the Aligned Rank Transform Contrast (ART-C) algorithm[29] (or alternatively for robustness Mann-Whitney-Wilcoxon) are reported, adjusted for the multiple comparisons with the Benjamini-Yekutieli method[30].

Heart rate data Figure 4, displays the *Low Frequency* (LF)/ *High Frequency* (HF) ratio. It is interpreted as an index of autonomic balance, where higher values indicate greater sympathetic dominance and lower values indicate greater parasympathetic dominance[36]. Panel **a** shows a significant difference between the *status-quo* and *curbless* scenarios (n=266, $p_{adj\ BY} = 1.902e-02$). Moreover, panel **c** reveals a significant difference in the LF/HF ratio between treatments: *parallel* viewers had a significantly higher ratio than *sequential* viewers (n=266, $p_{adj\ BY} = 9.925e-03$). Specifically, panel **b** shows how *parallel* viewers had a significantly higher heart rate when *approaching* the crossing than *sequential* viewers (n=266,

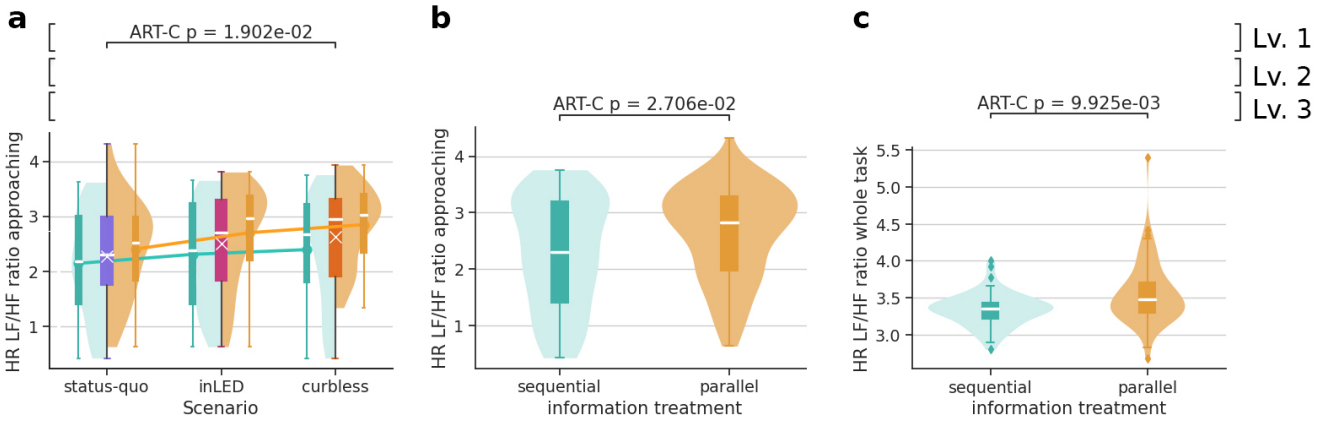

Figure 4. Physiological responses. Each panel presents statistically significant differences ($p < 0.05$) of low-frequency/high-frequency (LF/HF) ratio metric based on the heart-rate data for different subtasks within the VR session task: subtask *approaching* in panels **a**–**b** and whole task in panel **c**). When shown, scenarios are grouped on the X-axis, featuring the aggregated distribution as a coloured (blue, magenta and bright orange) box plot. Mean values are denoted with a white $\times$. Violin plots show distributions for each information treatment within the same scenario: mint for *sequential* and light orange for *parallel*. They are complemented by corresponding box plots displaying median η values and interquartile ranges. Lines linking information treatment boxplots show mean values visualizing within-treatment changes between scenarios. Annotations[26] on top of each panel display the contrast multifactor pairwise posthoc tests for the three levels of analysis after performing an Aligned Rank Transform (ART) ANOVA test[27]. Only significant results ($p < 0.05$) resulting from using the Aligned Rank Transform Contrast (ART-C) algorithm[29] are reported, adjusted for the multiple comparisons with the Benjamini-Yekutieli method[30].

$p_{adj\ BY} = 2.706e-02$).

Open-Answer Questions The analysis of open-answer questions involved a two-stage process. Firstly, all responses were read and labels were generated to summarise the answers in a meaningful way. Secondly, the labels were assigned using the LightTag software to ensure reproducibility.

The participants provided feedback on the *advantages and disadvantages* of the different scenarios for street crossing. Regarding the *status-quo* scenario, participants commonly noted its advantages such as its safety (n=19), familiarity (n=14), and ease of understanding (n=10). However, the most commonly mentioned disadvantage was the wait times for pedestrians and the fact that there is only one designated spot for crossing the street (n=15). The *inLED* scenario garnered mixed reviews from the participants. Among the advantages, the absence of a need to search for a crossing path for a long time was frequently noted (n=10). Participants also appreciated the clear and visible markings of the crossing, particularly at night (n=14). Additionally, a small number of participants found the scenario aesthetically pleasing (n=4). However, only a minority felt that the scenario was safe (n=5). In contrast, the most commonly mentioned disadvantage was the perceived complexity of the scenario (n=14). Participants also found the setting unfamiliar, which could be a potential hurdle to widespread adoption (n=8). Interestingly, a considerable number of participants expressed concern about the expensive cost of implementing and maintaining the inLED scenario (n=6). The *curbless* scenario generated mixed feedback from the participants. Many participants noted the convenience of being able to cross the street whenever they wanted (n=21), while a few found the scenario cost-efficient (n=4). However, the overwhelming majority of participants did not feel safe when crossing the street (n=24). Some participants assumed that the absence of visual cues meant that car drivers had no obligation to stop for pedestrians (n=8). Furthermore, a few participants found the scenario to be complicated and tiring (n=6). Interestingly, some participants also reported that cars frequently stopped too close to them while they were crossing the street, which was a notable disadvantage (n=3).

Participants were asked for their willingness to change their mode of transportation in different scenarios, and a range of responses was given. The *curbless* scenario received mixed feedback from the participants. Interestingly, the most common response was that participants did not want to be either a pedestrian (n=8) or a car driver (n=8) in the scenario. This might be due to the perception that scenario *curbless* had an increased probability that cars might hit someone. Moreover, some participants specifically stated a desire to switch from being a pedestrian to being in a car (n=6). However, a few participants found being a pedestrian in this scenario to be satisfactory (n=5). Regarding the *inLED* scenario, a majority of participants stated that they were comfortable with being a pedestrian in the scenario (n=7). Similarly, for the *status-quo* scenario, a comparable number of participants reported being fine with being a pedestrian (n=8).

During the study, participants were asked to identify the differences they observed while crossing the street in different scenarios. The majority of participants noted that the crosswalk and traffic lights were the main differences they noticed (n=24). Notably, a significant number of participants identified safety as the primary variable that varied across scenarios (n=12). Additionally, some participants believed that the primary difference was in the level of difficulty (n=4) or familiarity (n=4), which was an interesting observation.

Discussion

This study utilized immersive VR to evaluate the feasibility of three distinct street design scenarios that could be implemented in the near future and the impact of communication strategies to inform about them. The scenarios evaluated were: *status-quo*, *inLED*, and *curbless*. Potential benefits of these layouts were highlighted through a different allocation of street space by time-of-day, with public uses (e.g. street market, food stands, seating), but also the allocation of space to parking. Half of the participants viewed three videos depicting the design scenarios sequentially, whereas half of the participants viewed the videos simultaneously in split-screen, prior to being immersed in VR. Following the immersion in VR, participants answered a series of questions related to safety and perceived stress, as well as provided open-ended responses. Throughout the experiment, we recorded physiological (heart rate) data and tracked participants' behaviour through head movements and displacement while immersed in VR.

Our primary finding is that unconventional street layouts are less preferred and perceived as less safe in both the structured survey and open-answer questions. These results are mirrored by the heart rate data: We observe greater stress response in the *curbless* scenario, as it is unfamiliar, perceived as dangerous, and requires more cognitive resources to navigate.

The *inLED* and *curbless* are two extreme examples of shared space. Although there is no clear definition of shared space, all definitions agree that shared space is characterized by the removal of a clear separation between different road users[37], and by integrating vehicular traffic into the social space[6]. Research on shared space has resulted in mixed results, varying from pedestrians not minding to share space[38], to pedestrians preferring dedicated facilities[39]. From a driver's perspective, shared space results in anxiety and unease to drivers, resulting in possible lower speeds and more attention being paid to other shared space users[40].

In our case, a *status-quo* street design is preferred over unconventional layouts. Nevertheless, in-street technology, such as presented in the *inLED* scenario, is preferred over the *curbless* scenario. Such technology can demarcate pedestrian-only facilities and allows for flexible and adaptive use of space by time-of-day and day-of-week.

The *open-answer questions* shed more light on the perceived differences between the different scenarios. A frequently mentioned concern was safety, which aligns with the results of the structured question. The frequent mentioning of familiarity by participants suggests the significant cognitive effort required to adapt to new traffic paradigms and the strong preference for the status quo, which could potentially outweigh the benefits of adopting new transportation innovations due to the effects of the status quo bias[41]. While the safety distance to stop did not vary between the scenarios, the stopping distance appeared closer for the participants. This insight is valuable for planners, who need to take into account the effects of removing visual cues in street design.

The *inLED* scenario took the longest to complete, which is consistent with the open-answer questions where participants mentioned confusion due to the overwhelming information displayed. Interestingly, the *curbless* scenario without a timing mechanism took the longest to cross. This finding suggests that the convenience of the crossing can be increased by giving more time to pedestrians.

One indirect measure of presence in VR is head movement; increased head movement indicates that individuals are likely to behave and judge as in reality[42]. In two scenarios head movement was recorded. Our study found that the scenario influenced participants' visual exploration patterns with the *inLED* scenario requiring more focus on the ground plane and participants turning their heads from the left to right (head yaw), and looking more up-and-down (pitch) in the *status-quo* scenario. This result is intuitive, as the novel aspect of this scenario was the hexagonal road pavers that needed to be explored, directing more attention to the left and right rather than up and down. Exploratory research furthermore indicated that an increased head yaw, or scanning, might be due to higher anxiety[43]. Lower self-reported safety for the *inLED* scenario also points to higher anxiety in this scenario.

The content analysis of the answers given to the *open-answer questions* delivered further intriguing insights. Participants commented on the diverging maintenance costs for the different scenarios, suggesting a perception of their role not only as road users but also as citizens. This indicates the importance of not only highlighting the immediate benefits and drawbacks for road users but also organizational issues such as costs.

When asked about their willingness to change their mode of transportation in different scenarios, participants provided surprisingly mixed responses. In the *curbless* scenario, many participants were afraid of hitting pedestrians as car drivers and therefore explicitly stated not wanting to drive a car. On the contrary, other participants expressed fears of being hit by a car as pedestrians and therefore prefer to be a car driver. These findings highlight the importance of planners to consider the

unintended consequences of their designs on the behaviour of road users if the safety and equity of road use shall be ultimately improved.

When asked to state the differences, many participants identified safety, familiarity, and difficulty as key variables that varied across scenarios. This suggests that participants were more focused on the affordance of the space[44] and the behavioural effects of the new installations than the physical features. This highlights the importance of considering how citizens use and feel about a space when designing transportation infrastructure.

In our experiments, *parallel* viewers tended to favour conventional scenarios over unconventional ones more frequently than the sequential viewers. This finding indicates that having the ability to compare three traffic scenarios enabled participants to focus primarily on the drawbacks (instead of the advantages) of a novel scenario compared to a familiar one. It is likely not coincidental that the presentation of information made the greatest difference in questions relating to the vulnerable traffic participants, namely a child, and the transport mode with the greatest potential for harm: a car. Furthermore, our study found that *parallel* viewers needed longer to approach to cross, implying greater caution than *sequential* viewers. Furthermore, the information treatment appears to influence the comprehension of the scenarios. Specifically, *parallel* viewers showed a more constant movement pace, implying they may have comprehended the single scenario better than *sequential* viewers. These findings are mirrored by heart rate data: *parallel* viewing caused a shift toward sympathetic dominance which is frequently associated with stress or alert states.

Overall, our findings highlight the importance of seeking early feedback in any urban planning action and aligning with current practices of incremental and tactical urban planning. Participants' assumptions that the absence of visual cues influences car driver behaviour emphasize the importance of street design elements (e.g. markings, signs, delineated areas), to ensure driver awareness of their obligations to stop for pedestrians. In conclusion, our study has provided valuable insights into the impact of unconventional street layouts on road user behaviour and perceptions. Our findings suggest that such layouts are generally perceived as less attractive, less safe, and less preferred by participants. This highlights the importance of designing streetscapes that meets community needs and preferences. We also found that the ability to compare multiple scenarios allowed participants to more easily identify the drawbacks of a novel scenario compared to a familiar one. Physiological data supported our qualitative findings, suggesting increased perceived and measurable stress in unfamiliar scenarios. We recommend that planners and policy-makers consider multiple dimensions of street design innovations, including organizational issues such as costs and the immediate benefits and drawbacks for citizens. Additionally, we urge planners to take into account citizen perception to prevent unintended consequences such as the switch to less sustainable but apparently safer modes of transportation. Furthermore, we highlight the importance of tangible design elements, such as signs and demarcated areas, which increase perceived safety. We also recommend retaining simple markings as they change the perception of the space towards the positive. Finally, our study shows that participants were more focused on the consequences of changes and the behavioural effects of new installations rather than objective differences. This underscores the need for a user-centred approach when designing streets to meet the community's needs and preferences.

Materials and Methods

Experiment Design The experiment was conducted in collaboration with the Decision Science Laboratory at ETH Zurich. The laboratory was responsible for participant recruitment and payment, ensuring that experimental data and personal identifiable information were kept separate and safe by design. The study was preregistered[45] and received IRB approval from the Ethics Commission of ETH Zurich (IRB Approval Date: 2021-01-28, IRB Approval Number: EK 2020-N-183). All experiments were performed in accordance with applicable guidelines and regulations. All participants provided written informed consent before they participated in the experiment and were compensated for their time according to the laboratory's payment scheme.

The participant was briefed to cross the street in three different scenarios. Through this, a participant would experience the street from different perspectives. This enabled the participants to answer a series of questions on each scenario.

To that end, we designed three different design scenarios (Figure 5) tested in our experiments: In the *status-quo* scenario, the street was characterized by regular sidewalks, and a traffic light controlled the pedestrian crossing. The *inLED* scenario was characterized by the presence of a smart pavement consisting of hexagonal modular pavers[46,47] incorporating LEDs capable of changing colour based on the intended use of the street. In contrast, the *curbless* scenario relied on smart vehicle sensing and the adaptive behaviour of people.

Apart from the differences in designs, all three scenarios were designed similarly (Figure 6). Each scenario featured a park adjacent to a straight long road with curves at both ends of the road. The curves blocked the line of sight and allowed for the generation of pedestrians and cars out of sight of the participant. The participant could freely move in this area but was discouraged from moving too far in one direction along one side of the street, namely by limiting teleportation to certain areas (i.e. alongside the street) and potentially by the instructors reminding them that their task was to cross the street.

Each of the three street scenarios was experienced twice in two different configurations by all the participants (i.e., within-subject factor), once during rush hour and once during the evening. The scenarios were designed such that street usage varied

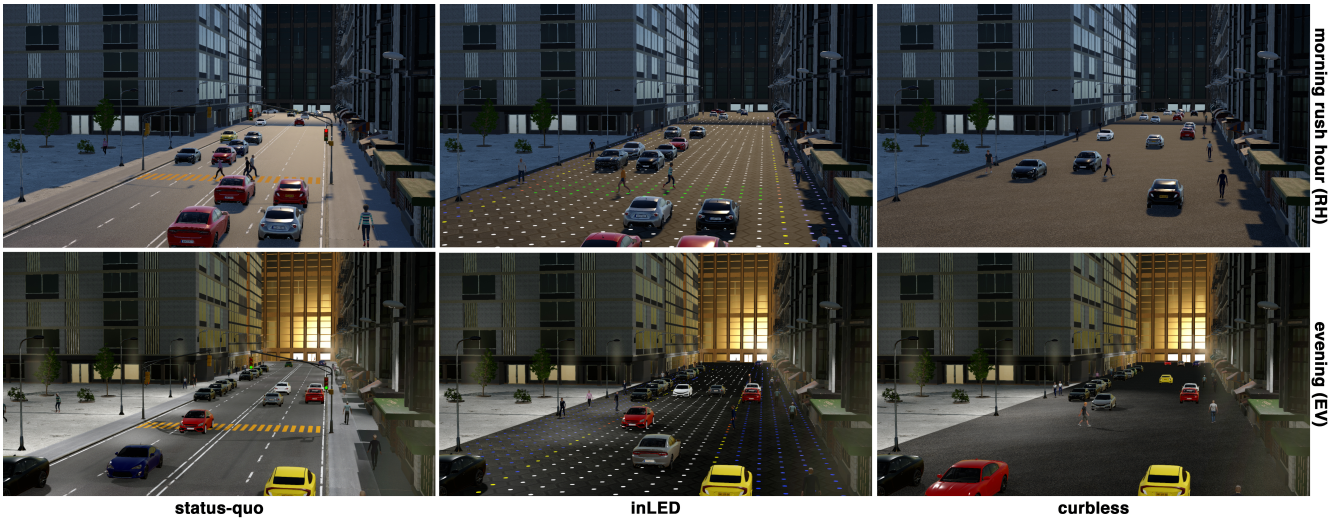

Figure 5. Participants crossed a street wearing VR headsets in three different design scenarios: conventional traffic light controlled pedestrian crossing (*status-quo*, left), dynamic pavement for right-of-way allocation (*inLED*, centre), and flat street (*curbless*, right). Additionally, each traffic scenario was presented for two different times of the day with distinctive lighting, use, and traffic conditions: morning rush hour (top row) and evening traffic (bottom row). (For more details, see Table 2).

throughout the day. During rush hour, the street was characterized by a steady stream of commuters, while in the evening, it was used by families as a space for play and dining. Only the order of the scenarios was randomized across participants.

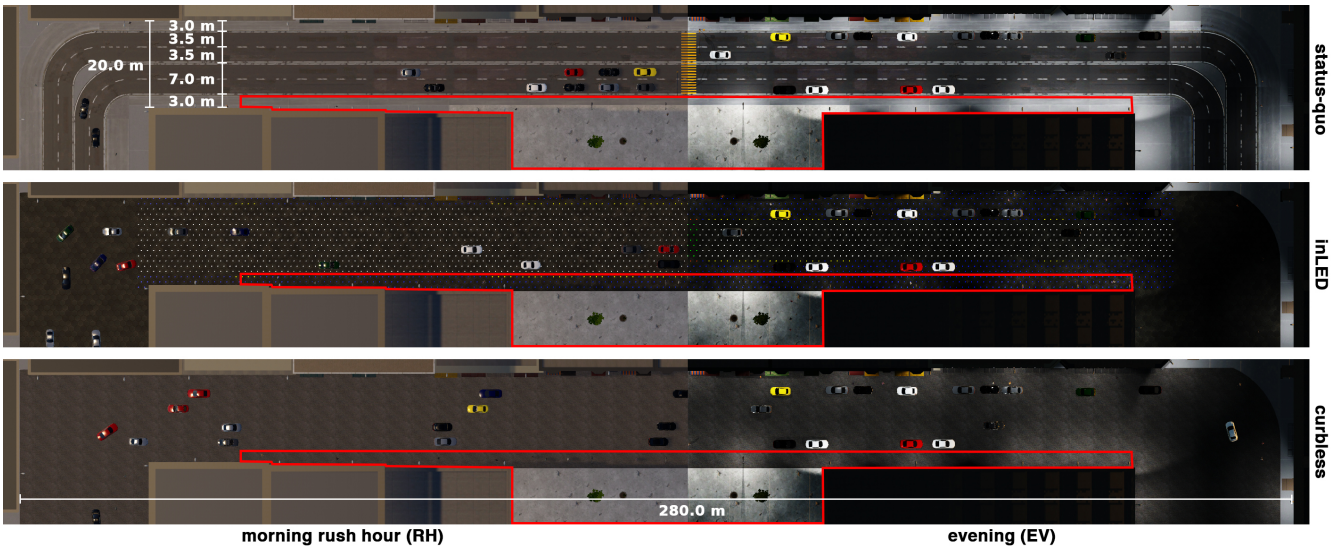

Figure 6. Floor plans of the three designed VR scenarios (*status-quo*, *inLED*, and *curbless*) and the two configurations (*morning rush hour*, and *evening*). In red is highlighted the area where teleporting was allowed (For more details, see Table 2).

Prior to starting the VR task, participants filled out a SSS for establishing a stress level baseline for after-treatment differences assessment (see **a** in Figure 1). Then, participants watched a 3-minutes video to become familiar with the traffic environment. The presentation of the videos varied between subjects: Participants in the *sequential* group watched videos from a bird's-eye view for every scenario separately in sequential order matching the randomized order experienced afterwards for performing the task in the three VR scenarios. Instead, participants in the *parallel* group watched the same videos, but on a split-screen, simultaneously in a side-by-side layout, allowing for a direct comparison across scenarios. This gave them additional qualitative information about each street design. In both cases, the window for the video clip of each scenario was depicted in the same size and aspect ratio: i.e., each scenario video window occupied 1/3 of the width of the screen, resulting in *sequential* screen layout having one single video playing with most of the screen empty, and *parallel* visualization having

all the three videos filling the entire width of the screen (see **b** in Figure 1). In Table 2 a summary of the treatments for the experiment design is shown with a motivation of the interventions, both for the between-subjects information treatments and for the within-subjects scenarios.

The task that participants had to perform was identical in every design scenario and configuration (see **d** in Figure 1). Moreover, they had two attempts to perform the task per configuration in case of being hit by a car and failing. To complete the task, participants had three movement options, which could be freely switched between (1) physical walking and turning, (2) virtual walking and turning using a controller, and (3) teleportation. Teleportation was only possible on the sidewalk and restricted in range to 8 meters to prevent participants from simply teleporting across the street. The virtual walking and teleportation options were introduced due to the large space requirements of the VR setting, although participants also had ample physical space to move around in a 10m×5m area. After ensuring a comfortable fit of the head-mounted display, participants were given two controllers and found themselves in a virtual tutorial environment. An instructor guided them through a fixed tutorial sequence for getting accustomed to VR and learning the different locomotion techniques (see **f** in Figure 1).

After completing the task, in both configurations for each design scenario, participants were asked to fill in a survey assessing their emotional state and perception of the scenario (see **e** in Figure 1). To complete the survey, they removed the VR headset and switched to a desktop computer. The survey included questions related to stress, perceived safety, and willingness to cross and share the street. At the very end of the experiment, after performing the task in the three scenarios, participants were asked to reply to a set of open-answer questions regarding general impressions and behaviour change based on the scenarios, as well as general socio-demographic data. All these questions that were answered are listed in the paragraph titled "questionnaires."

To gain a deeper understanding of how street design influences the walking experience of pedestrians[48,49], we captured participants' physiological responses. Specifically, we measured participants' heart rate and heart rate variability, as well as their activity levels, using a breast strap (Garmin HRM-Pro).

For conducting the VR part of the experiment, our setup consisted of a Meta Quest 2 head-mounted display connected to a desktop computer (Intel® Core™ i7-11700 11th generation [8 Cores, 16 MB Cache, 2,5 GHz up to 4,9 GHz], NVIDIA® GeForce® RTX 3070™ 8 GB GDDR6; 32 GB [16 GB x 2, DDR4, 2.933 MHz]; M.2-PCIe-SSD with 1 TB + SATA with 1 TB, 7200 1/min) via Air Link. The VR application was developed in Unity 2020.3.32f1 using the C# programming language. The virtual environment depicted in Figure 5 was designed with assets from the Unity Asset Store. Realistic, rigged 3D models of pedestrians were taken from Mixamo[50].

Data collection and pre-processing Each participant (n=43) went through each scenario (*status-quo*, *inLED*, *curbless*) at least twice, one per configuration (morning rush hour, evening). Additional attempts resulted from second opportunities in case of failure to perform the task.

Physiological measurements were taken continuously from the beginning of the experiment until the end resulting in a continuous time series (more details in paragraph *Heart Rate data*). Behavioural data were taken separately at every VR session, defined as every run to perform the task in a given scenario and configuration until the participant accomplished crossing the street or failed because of being hit by a car ($n_{sessions} = 266$). Behavioural data were computed from the VR environment logs, which recorded the position (x, y, z) and orientation of the VR headset in three angles (*pitch*, *yaw*, *roll*). From this raw log data, *distance walked*, *speed*, and angle change for *pitch*, *yaw*, *roll*, and for*total angle* (as the sum of *pitch* and *yaw*) were computed. The structured self-reported survey (comprising SSS, acceptance, safety perception and assessment of willingness to cross) was answered after completing an entire scenario in both configurations ($n_{structured\ surveys} = 129$).

Physiological and behavioural time series were synchronized and matched. Later physiological continuous data was cropped according to the starting and ending time of each session to have a combined time series per VR session including physiological and behavioural data.

We recorded the first-person perspective in VR of participants for the entire duration of the experimental session. These recordings were later manually timestamped, which allowed mapping all data points to four different subtasks. The first subtask, which involved participants exploring the virtual environment, was labelled as *familiarization* (1). Following this, participants typically walked towards the curb (or perceived border of the area suitable for cars) in preparation for the road-crossing task, which we referred to as *approaching* (2). Subsequently, each participant had to wait in front of a (virtual) crosswalk, a subtask we termed *waiting* (3) until the red light turned green or he felt it was safe to cross. Finally, participants completed the task by crossing the street, a subtask we called *crossing* (4). The subdivision of the task into these subtasks facilitated comparisons of completion times across the different stages as well as highlighting particular differences within the task. Consequently, time series are cropped as well to match these subtasks. In the case of the lack of one of the subtasks (e.g., because a participant didn't wait before starting to walk or before crossing), missing values are imputed with the closest 10 seconds of data from the contiguous subtasks. Time series for each session are summarised by computing their means (ν), medians (η), and standard deviations (σ). These values were later used for statistical analysis.

Table 2. Characterization of scenarios and motivations.

		name	description	motivations / aims
information treatments	randomized	*sequential*	Each video only shows one scenario at a time. Each scenario is displayed for 1 min. Scenarios are displayed consecutively (one after another). Each scenario is displayed once. Size of video window is 1/3 of the treatment *parallel.* Total visualization time of all videos is 3 min.	Baseline for explanation of scenarios.
		parallel	The 3 scenarios are displayed simultaneously. Side-by-side window layout. Each loop (with the 3 scenarios) lasts 1 min. The loop with the 3 scenarios is repeated 3 times. Window for each scenario has same size as in *sequential.* Total visualization time of videos is 3 min.	Enhancing comparison between scenarios. Highlighting advantages and disadvantages. Synoptic presentation.
Scenarios	randomized	*status-quo*	Conventional marking and lanes. Conventional segregation of roadways and sidewalks. Regular crossing with traffic light control.	Traffic level of service. Safety. Maximize max speeds and throughput.
		inLED	Flat street. No curbs, no hard barriers, no hard separations. Dynamic pavement lighting to sign changing uses and rights of way.	Flexible use of streets. Maximizing available space. Dynamic setting. Minimize speed variations. Keep constant speed. Matching effective average speed. Chasing slower is faster effects. Environmental improvement. Enhancing IoT and sensing. Relies on hard infrastructure
		curbless	Flat street. No curbs, no hard barriers, no hard separations. No marking lights. Conventional pavement.	Easy to maintain and implement. Low infrastructure intervention. Low cost. Relies on autonomous driving.
Configurations	not randomized	*morning rush hour (RH)*	High level of car traffic. 2 directions of traffic. 2 lanes per direction. High level of pedestrian traffic.	Increase diversity of urban situations. Diversify stimuli and visual noise. Cause more diversity of responses. This variation is not randomized, the experiment does not control for this factor.
		evening (EV)	Higher level of diversity of uses in the street. 2 directions of traffic. 1 lane per direction. Street parking band combined. Higher level of pedestrian traffic. More visual noise. Darker environment with artificial light.	

The answers to the structured self-reported survey were normalized for each participant by computing the differences between each value reported after experiencing a scenario and the initial baseline value registered by each participant at the beginning of the experiment. These differences are later used in the statistical tests.

Heart Rate Data Participants in VR environments are exposed to a variety of stimuli that can elicit a variety of physiological and psychological responses[51]. A variety of sensors, including heart rate sensors, head movement sensors, and displacement sensors, can be used to measure these responses. Heart rate variability (HRV) is a measurement of the time difference between successive heartbeats. It is a useful indicator of the body's physiological response to stress and other stimuli. Heart rate sensors can be used in VR experiments to track participants' HRV as they navigate virtual environments. This data can reveal how participants react to various aspects of the environment, such as visual cues or sound effects.

We collected heart rate data continuously from the start to the end of the VR experiment. Outliers in the dataset were removed to guarantee data quality. In such an experimental setting, it is possible to have outliers in the collected heart rate data. Outliers can be caused by a variety of factors, including measurement errors, physiological abnormalities, and participant noncompliance. While pre-processing the data, we looked at the outliers and removed them in a way that didn't skew the findings. We resampled and adjusted the data to one sample per second. Timestamps between different sensing sources, including the VR environment logs and sensing devices, were matched to accurately align data across sources. The whole heart rate time series for each participant was split into individual sessions corresponding to each VR scenario.

In VR experiments, analyzing the Power Spectral Density (PSD) of physiological signals such as heart rate can provide insights into the autonomic nervous system response, which can help in understanding the participants' emotional or cognitive states[52,53]. The PSD represents the power of various frequencies in a signal mathematically. The PSD gives information about the distribution of power at different frequencies in the heart rate signal in the case of heart rate variability. The PSD's HF component is linked to parasympathetic (vagal) activity, whereas the PSD's LF component is linked to sympathetic and parasympathetic activity[54].

We analyzed the PSD of heart rate data from various participants to investigate the effects of diverse scenarios in a VR experiment on the autonomic nervous system. We used a wearable device (Garmin HRM-Pro) to collect heart rate data, and each participant went through a series of VR scenarios, each intended to elicit a particular emotional response. We used the Welch method to analyze PSD because it reduces noise and gives a more accurate estimate of the PSD[55]. To obtain a more accurate estimate of the PSD, the Welch technique divides the signal into overlapping segments and then averages the periodograms of these segments. The technique reduces noise caused by the signal's non-stationarity by using overlapping segments, and it improves the estimate of the PSD by averaging the periodograms. The PSD was then integrated in the LF (0.04-0.15 Hz) and HF (0.15-0.4 Hz) bands to determine the total power in each band. The LF/HF ratio was determined as the ratio of LF power to HF power. The LF/HF ratio is a commonly used metric for autonomic nervous system activity. A higher LF/HF ratio is commonly read as a sign of sympathetic dominance or a shift in the balance of sympathetic and parasympathetic activity toward sympathetic dominance. This could be due to stress, nervousness, or other conditions that cause sympathetic activation. A lower LF/HF ratio, on the other hand, is frequently linked with parasympathetic dominance or a shift towards parasympathetic activity, which may be associated with relaxation and recovery[36].

Statistical Details To characterise the independent and dependent variables of the experiment, we employed the Shapiro-Wilk test[56] to check normality and the Levene test[57] for equal variance. Different statistical methods were applied depending on the type and distribution of the data for assessing the association between variables and the informational treatments and scenarios. However, as the majority of the experiment variables did not follow a normal distribution, we utilized non-parametric tests to evaluate differences across treatments and scenarios.

For categorical variables, both independent (i.e., gender, experience with VR experiments, ownership of a car, driving license, and walking habits,) and dependent (i.e., counts of failed attempts for crossing the street, favourite scenario reported at the end of each experiment) a two-tailed Fisher's exact test[58] was used on the contingency table of occurrences for each treatment group due to the small sample size of some of them (i.e., below 5).

We statistically tested the independence of the treatments. The only variables from the socio-demographic data that were not completely randomized and showed unbalance with regards to the treatment groups were ownership of a car (n=43, $p_{two-sided\ Fisher} = 1e-03$), driving license ownership (n=43, $p_{two-sided\ Fisher} = 2.499e-02$), and reported previous experience in VR experiments (n=43, $p_{two-sided\ Fisher} = 5e-04$). Hence, we controlled for the effect of these covariates in combination with the main interventions (*information treatment*, and *scenario*) in all the statistical analyses and they didn't show any significant effect on the responses ($p > 0.05$).

For the analysis of the effects of the interventions, we performed an Aligned Rank Transform (ART) ANOVA test. This allowed us to perform factorial non-parametric analyses when violations in the assumptions of normality and variance were present and it could handle repeated measures[27] as in the case of the mixed experiment designed.

We tested for the separated effect of each intervention ($Response \sim Information\ treatment + (1|Participant)$ and $Response \sim$

$Scenario + (1|Participant)$) and the interaction between them ($Response \sim Information\ treatment \times Scenario + (1|Participant)$). Given the repeated measures design of the experiment, the main interventions (*information treatment* and *scenario*) were considered as fixed effects, and the repeated measures on the participants on consecutive sessions across scenarios were included as a random effect ($(1|Participant)$).

In those response variables where the random effect for each participant was not significant, an alternative non-parametric Kruskal-Wallis H test for the analysis of variance[28] was performed for robustness, whose results are reported and shown in figures. No differences regarding the significance of the main effects were found between both procedures.

The sample size (n) for the omnibus analysis of variance changed depending on the response variables and the number of repeated measures taken over the number of participants (n=43). Dependent variables resulting from the structured self-reported survey were recorded every time after each scenario (n=129). Differently, behavioural and physiological responses were taken during each session (n=266).

In the case of statistical significance ($p < 0.05$) of the omnibus test for variance, it follows a posthoc test to determine the statistical significance of their differences. In the case of using the ART procedure, we performed a multifactor contrast test at different levels using the algorithm ART-C, which is an extension of the former to reduce the Type I error rated with higher statistical power than the t-test, Mann-Whitney U test, Mann-Whitney-Wilcoxon test, and ART test, particularly for nonconforming data[29]. Alternatively, for the posthoc test after the Kruskal-Wallis test, we used a two-sided Mann-Whitney-Wilcoxon test[59]. Finally, p-values are adjusted for the multiple comparisons using the Benjamini-Yekutieli (BY) method[30]. This adjustment controls Type I error of multiple comparisons based on the *false discovery rate* (FDR), which improves the statistical power when a large number of comparisons are simultaneously tested[60]. BY is a conservative test and compromises slightly the discovery rate by not making assumptions on the dependency of hypotheses.

Finally, in some cases, after performing a detailed assessment of the distribution of response variables in *information treatment* groups within *scenarios* (i.e., interactions *Information treatment* $\times$ *Scenario*), for the within-scenario pairwise contrast tests (level 3), a one-sided Mann-Whitney-Wilcoxon test can be used, with a significance level $p < 0.1$, for its higher power due to the particular characteristics of the compared distributions which can allow making a strong hypothesis about the values[31,61].

Questionnaires

Answers to the following questions were recorded using Qualtrics.

Open-answer questionnaire Participants were provided with an open text field, where they could write as much as they wanted in response to the following questions:

- In your opinion, what are the main advantages of each of the three scenarios? What are the main disadvantages?
- What are the most noticeable differences between the three scenarios for you?
- For each or any scenario, would you want to change your usual mode of transportation and why?

Socio-demographic questionnaire The socio-demographic questionnaire was answered at the end of the experiment.

- In which country/city do you live? (dropdown menu w/ list of countries)
- In which city were you raised? (dropdown menu w/ list of cities)
- How do you define your neighborhood? (city center/downtown, compact walkable neighborhood, low density suburb, rural area/village)
- What is your year of birth? (open field, only NATURAL numbers allowed)
- I identify as (male, female, non-binary/third gender, prefer not to say)
- What is your subject of study/profession? (dropdown menu with ETH specific choices)
- How often did you take part in experiments? (0, 1–3, >4)
- Do you have experience with VR? (Y/N)
- Do you have a driving license? (Y/N)
- Do you own a car? (Y/N)
- How many kilometers do you drive per week on average (0, <10, 10–50, 50–100, >100).

- What is your preferred mode of transportation for your daily chores (work, groceries, etc.)? (walking, bike, public transportation, motorbike, car)
- Do you walk (>500 m) daily? (Y/N)

Short Stress State (SSS) Questionnaire
All of the items of the short stress state questionnaire[32] are answered on a 10-point scale, once before and subsequently after each scenario.

- I feel dissatisfied.
- I feel alert.
- I feel depressed.
- I feel sad.
- I feel active.
- I feel impatient.
- I feel annoyed.
- I feel angry.
- I feel irritated.
- I feel grouchy.

Safety, willingness, attractiveness In addition to the short stress state questionnaire, the following questions were asked after participants encountered each street scenario:

- How safe do you feel due to the proximity and number of pedestrians/cars? (1–7)
- How willing are you to cross the street with a kid/at night? (1–7)
- Please rate how attractive you found the scenario as a pedestrian. (0–5)

Additional information

Acknowledgements JASV, CIH, SM, MM, MS and DH acknowledge support through the project "CoCi: Co-Evolving City Life", which has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No. 833168.

The authors would like to express their gratitude to several individuals who provided valuable assistance and support during this study. We thank Stefan Wehrli for his expert feedback on experimental methods, as well as for his invaluable assistance in carrying out the experiments. We are also grateful to Michal Gath Morad for providing insightful comments during a productive zoom call. Additionally, we extend our thanks to Fabian Schläfli for his thoughtful feedback on the project during a helpful zoom call. Their contributions were instrumental in the successful completion of this research, and we deeply appreciate their time, effort, and expertise contributed.

Author contributions statement JASV proposed and designed the three scenarios and co-designed the virtual traffic situations. He supervised the students, especially in drafting the introduction, and contributed to writing the manuscript, and the figures. CIH worked out the details of the treatment variations and supervised the students. She co-developed ideas for scenarios and situations, drafted the hypotheses, and introduced the idea of using sensors to monitor physiological data. Furthermore, she specified procedural details with the Decision Science Laboratory (DeSciL) and acquired ethics approval. She also contributed to writing the manuscript. SM contributed technical details on responsive street technology and searched for appropriate sensors as well as papers describing a connection between sensor data and behaviour. He performed heart rate data analysis and also contributed to writing the manuscript. MS and MM developed the virtual reality (VR) setting and contributed to writing the manuscript. DH proposed to study adaptive infrastructures and flexible uses of streets, acquired funding, engaged the VR developers, provided feedback and ideas, and contributed to the manuscript. MvE read the paper and provided feedback on experimental details. He also contributed a section on pedestrian-vehicle interaction. All authors critically reviewed and edited the manuscript and approved the final version for submission.

Data availability statement To ensure transparency and reproducibility of our research, we will make raw and processed data, analysis scripts and code publicly available after acceptance via an open access Github repository. The working repository, https://github.com/ethz-coss/VR_future_streets, can be accessed upon invitation.

Competing interests The authors declare no competing interests.

References

1. Angel, S. *et al. Atlas of Urban Expansion - Volume 2: Block and Roads* (2016), 2016 edn.
2. van Doorn, L., Amanprit, A. & Rapoport, E. Hot property, DOI: 10.1007/978-3-030-11674-3 (2019).
3. ITF. Streets that fit re-allocating space for better cities (2022).
4. von Schönfeld, K. C. & Bertolini, L. Urban streets: Epitomes of planning challenges and opportunities at the interface of public space and mobility. *Cities* **68**, 48–55, DOI: 10.1016/j.cities.2017.04.012 (2017).
5. Valença, G., Moura, F. & de Sá, A. M. Main challenges and opportunities to dynamic road space allocation: From static to dynamic urban designs. *J. Urban Mobil.* **1**, 100008, DOI: 10.1016/j.urbmob.2021.100008 (2021).
6. Hamilton-Baillie, B. Shared space: Reconciling people, places and traffic. *Environment* **34**, 161–181 (2008).
7. Meyboom, A. *Driverless Urban Futures* (Routledge, 2018).
8. Schlossberg, M., Riggs, W., Millard-Ball, A. & Shay, E. Rethinking the Street in an Era of Driverless Cars. Tech. Rep., University of Oregon, APRU, Sustainable Cities Initiative (2018).
9. Papageorgiou, M., Mountakis, K. S., Karafyllis, I., Papamichail, I. & Wang, Y. Lane-Free Artificial-Fluid Concept for Vehicular Traffic. *Proc. IEEE* **109**, 114–121, DOI: 10.1109/JPROC.2020.3042681 (2021). 1905.11642.
10. Pérez-Méndez, D., Gershenson, C., Lárraga, M. E. & Mateos, J. L. Modeling adaptive reversible lanes: A cellular automata approach. *PLOS ONE* **16**, e0244326, DOI: 10.1371/journal.pone.0244326 (2021).
11. Kondor, D. *et al.* Addressing the "minimum parking" problem for on-demand mobility. *Nat. Sci. Reports* **10**, 1–10 (2020).
12. Rieland, R. Rethinking How We Build City Streets (2018).
13. Ruiz-Apilánez, B., Karimi, K., García-Camacha, I. & Martín, R. Shared space streets: Design, user perception and performance. *Urban Des. Int.* **22**, 267–284, DOI: 10.1057/s41289-016-0036-2 (2017).
14. Zacharias, J. Pedestrian behavior and perception in urban walking environments. *J. Plan. Lit.* **16**, 3–18, DOI: 10.1177/08854120122093249 (2001).
15. Duarte, F. & Ratti, C. The Impact of Autonomous Vehicles on Cities: A Review. *J. Urban Technol.* **25**, 3–18, DOI: 10.1080/10630732.2018.1493883 (2018).
16. Gössling, S. Why cities need to take road space from cars - and how this could be done. *J. Urban Des.* **25**, 443–448, DOI: 10.1080/13574809.2020.1727318 (2020).
17. Rosique, F., Navarro, P. J., Fernández, C. & Padilla, A. A systematic review of perception system and simulators for autonomous vehicles research, DOI: 10.3390/s19030648 (2019).
18. Gold, C., Körber, M., Hohenberger, C., Lechner, D. & Bengler, K. Trust in Automation – Before and After the Experience of Take-over Scenarios in a Highly Automated Vehicle. *Procedia Manuf.* **3**, 3025–3032, DOI: 10.1016/j.promfg.2015.07.847 (2015).
19. Jayaraman, S. K. *et al.* Pedestrian Trust in Automated Vehicles: Role of Traffic Signal and AV Driving Behavior. *Front. Robotics AI* **6**, 117, DOI: 10.3389/frobt.2019.00117 (2019).
20. Sobhani, A. & Farooq, B. Impact of smartphone distraction on pedestrians' crossing behaviour: An application of head-mounted immersive virtual reality. *Transp. Res. Part F: Traffic Psychol. Behav.* **58**, 228–241, DOI: 10.1016/j.trf.2018.06.020 (2018).
21. Nazemi, M. *et al.* Studying bicyclists' perceived level of safety using a bicycle simulator combined with immersive virtual reality. *Accid. Analysis & Prev.* **151**, 105943 (2021).
22. Drettakis, G., Roussou, M., Reche, A. & Tsingos, N. Design and evaluation of a real-world virtual environment for architecture and urban planning. *Presence* **16**, 318–332 (2007).
23. Bhagavathula, R., Williams, B., Owens, J. & Gibbons, R. The reality of virtual reality: A comparison of pedestrian behavior in real and virtual environments. vol. 3, 2056–2060, DOI: 10.1177/1541931218621464 (Human Factors and Ergonomics Society Inc., 2018).
24. Deb, S., Carruth, D. W., Sween, R., Strawderman, L. & Garrison, T. M. Efficacy of virtual reality in pedestrian safety research. *Appl. Ergonomics* **65**, 449–460, DOI: 10.1016/j.apergo.2017.03.007 (2017).
25. Farooq, B., Cherchi, E. & Sobhani, A. Virtual immersive reality for stated preference travel behavior experiments: A case study of autonomous vehicles on urban roads. *Transp. Res. Rec.* **2672**, 35–45, DOI: 10.1177/0361198118776810 (2018).

26. Charlier, F. *et al.* Statannotations, DOI: 10.5281/zenodo.7213391 (2022).

27. Wobbrock, J. O., Findlater, L., Gergle, D. & Higgins, J. J. The Aligned Rank Transform for nonparametric factorial analyses using only ANOVA procedures. In *Conference on Human Factors in Computing Systems - Proceedings*, 143–146, DOI: 10.1145/1978942.1978963 (Association for Computing Machinery, 2011).

28. Kruskal, W. H. & Wallis, W. A. Use of Ranks in One-Criterion Variance Analysis. *J. Am. Stat. Assoc.* **47**, 583–621, DOI: 10.1080/01621459.1952.10483441 (1952).

29. Elkin, L. A., Kay, M., Higgins, J. J. & Wobbrock, J. O. An Aligned Rank Transform Procedure for Multifactor Contrast Tests. In *The 34th Annual ACM Symposium on User Interface Software and Technology*, vol. 15, 754–768, DOI: 10.1145/3472749.3474784 (ACM, New York, NY, USA, 2021). 2102.11824.

30. Benjamini, Y. & Yekutieli, D. The control of the false discovery rate in multiple testing under dependency. *Annals Stat.* **29**, 1165–1188, DOI: 10.1214/aos/1013699998 (2001).

31. Fagerland, M. W. & Sandvik, L. The Wilcoxon-Mann-Whitney test under scrutiny. *Stat. Medicine* **28**, 1487–1497, DOI: 10.1002/sim.3561 (2009).

32. Helton, W. S. Validation of a short stress state questionnaire. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 48, 1238–1242 (Sage Publications Sage CA: Los Angeles, CA, 2004).

33. Bischof, W. F., Anderson, N. C., Doswell, M. T. & Kingstone, A. Visual exploration of omnidirectional panoramic scenes. *J. Vis.* **20**, 23–23, DOI: 10.1167/JOV.20.7.23 (2020).

34. van Helvoort, D., Stobbe, E., Benning, R., Otgaar, H. & van de Ven, V. Physical exploration of a virtual reality environment: Effects on spatiotemporal associative recognition of episodic memory. *Mem. Cogn.* **48**, 691–703, DOI: 10.3758/s13421-020-01024-6 (2020).

35. Hu, B., Johnson-Bey, I., Sharma, M. & Niebur, E. Head movements during visual exploration of natural images in virtual reality. In *2017 51st Annual Conference on Information Sciences and Systems, CISS 2017*, DOI: 10.1109/CISS.2017.7926138 (Institute of Electrical and Electronics Engineers Inc., 2017).

36. Shaffer, F. & Ginsberg, J. P. An overview of heart rate variability metrics and norms. *Front. public health* 258 (2017).

37. Karndacharuk, A., Wilson, D. J. & Dunn, R. A review of the evolution of shared (street) space concepts in urban environments. *Transp. Rev.* **34**, 190–220, DOI: 10.1080/01441647.2014.893038 (2014).

38. Hammond, V. & Musselwhite, C. The attitudes, perceptions and concerns of pedestrians and vulnerable road users to shared space: A case study from the uk. *J. Urban Des.* **18**, 78–97, DOI: 10.1080/13574809.2012.739549 (2013).

39. Kaparias, I., Bell, M. G., Miri, A., Chan, C. & Mount, B. Analysing the perceptions of pedestrians and drivers to shared space. *Transp. Res. Part F: Traffic Psychol. Behav.* **15**, 297–310, DOI: 10.1016/j.trf.2012.02.001 (2012).

40. Moody, S. & Melia, S. Shared space – research, policy and problems. *Proc. Inst. Civ. Eng. - Transp.* **167**, 384–392, DOI: 10.1680/tran.12.00047 (2014).

41. Samuelson, W. & Zeckhauser, R. Status quo bias in decision making. *J. risk uncertainty* **1**, 7–59 (1988).

42. Slater, M., Steed, A., McCarthy, J. & Maringelli, F. The influence of body movement on subjective presence in virtual environments. *Hum. Factors: The J. Hum. Factors Ergonomics Soc.* **40**, 469–477, DOI: 10.1518/001872098779591368 (1998).

43. Won, A. S., Perone, B., Friend, M. & Bailenson, J. N. Identifying anxiety through tracked head movements in a virtual classroom. *Cyberpsychology, Behav. Soc. Netw.* **19**, 380–387, DOI: 10.1089/cyber.2015.0326 (2016).

44. Gibson James, J. *The ecological approach to visual perception* (Psychology Press, 1979).

45. Argota Sánchez-Vaquerizo, J. & Hausladen, C. I. Citizen Perception of Future Street Scenarios, DOI: https://doi.org/10.1257/rct.10003-1.0 (2022).

46. Ratti, C. The Dynamic Street (2018).

47. De Larrard, F., Sedran, T. & Balay, J.-M. Removable urban pavements: an innovative, sustainable technology. *Int. J. Pavement Eng.* **14**, 1–11 (2013).

48. Coensel, B. D., Vanwetswinkel, S. & Botteldooren, D. Effects of natural sounds on the perception of road traffic noise. *The J. Acoust. Soc. Am.* **129**, EL148–EL153 (2011).

49. Wilkie, S. & Stavridou, A. Influence of environmental preference and environment type congruence on judgments of restoration potential. *Urban For. & Urban Green.* **12**, 163–170 (2013).

50. Adobe. Mixamo (2022).

51. Martens, M. A. *et al.* It feels real: physiological responses to a stressful virtual reality environment and its impact on working memory. *J. Psychopharmacol.* **33**, 1264–1273 (2019).

52. Kamath, M. V. & Fallen, E. L. Power spectral analysis of heart rate variability: a noninvasive signature of cardiac autonomic function. *Critical reviews biomedical engineering* **21**, 245–311 (1993).

53. Anderson, A. P. *et al.* Relaxation with immersive natural scenes presented using virtual reality. *Aerosp. medicine human performance* **88**, 520–526 (2017).

54. Montano, N. *et al.* Power spectrum analysis of heart rate variability to assess the changes in sympathovagal balance during graded orthostatic tilt. *Circulation* **90**, 1826–1831 (1994).

55. Welch, P. The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on audio electroacoustics* **15**, 70–73 (1967).

56. Shapiro, S. S. & Wilk, M. B. An Analysis of Variance Test for Normality (Complete Samples). *Biometrika* **52**, 591, DOI: 10.2307/2333709 (1965).

57. Conover, W. J., Johnson, M. E. & Johnson, M. M. A Comparative Study of Tests for Homogeneity of Variances, with Applications to the Outer Continental Shelf Bidding Data. *Technometrics* **23**, 351–361, DOI: 10.1080/00401706.1981.10487680 (1981).

58. Fisher, R. A. On the Interpretation of χ 2 from Contingency Tables, and the Calculation of P. *J. Royal Stat. Soc.* **85**, 87, DOI: 10.2307/2340521 (1922).

59. Mann, H. B. & Whitney, D. R. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals Math. Stat.* **18**, 50–60, DOI: 10.1214/aoms/1177730491 (1947).

60. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. Royal Stat. Soc. Ser. B (Methodological)* **57**, 289–300, DOI: 10.1111/j.2517-6161.1995.tb02031.x (1995).

61. Fuest, T., Schmidt, E. & Bengler, K. Comparison of methods to evaluate the influence of an automated vehicle's driving behavior on pedestrians: Wizard of Oz, virtual reality, and video. *Inf. (Switzerland)* **11**, 291, DOI: 10.3390/INFO11060291 (2020).