The identification of factors affecting drivers' perceived risk in pedestrian-vehicle interaction: A crowdsourcing study

Master Thesis

Bram Kooijman

TU Delft

The identification of factors affecting drivers' perceived risk in pedestrian-vehicle interaction: A crowdsourcing study

By

Bram Kooijman
(4751086)

in partial fulfilment of the requirements for the degree of

Master of Science
in BioMechanical Design

at the Delft University of Technology,
to be defended publicly on 13 October, 2021 at 13:00.

Supervisors:	Dr. ir. J. C. F. De Winter Dr. D. Dodou Dr. P. Bazilinskyy
Thesis committee:	Dr. ir. J. C. F. De Winter, Dr. D. Dodou, Dr. P. Bazilinskyy, Dr. ir. R. Happee

An electronic version of this thesis is available at *http://repository.tudelft.nl/*.

TU Delft

Acknowledgements

The past year was an unusual year. A year that was mainly directed by the pandemic, with high ups and low downs. Although the process of writing this thesis was very educational, it was also a very individual process, meaning much working from home, in which I had to stay vigilant for a decrease in energy and motivation. Nevertheless, an interesting subject, enthusiastic and helpful supervisors, equally depressed friends and an overconsumtion of coffee is what got me through the year and at the end of my time as a student.

First of all, I would like to thank my supervisors Pavlo Bazilinskyy, Joost de Winter and Dimitra Dodou. Pavlo helped me greatly with writing the python code, providing input through brainstorming sessions and offer the necessary feedback while keeping the enthusiasm up. Although Pavlo was my daily supervisor, Joost and Dimitra were equally involved and were always available for critical comments and interesting discussions. In conclusion, thank you for all the help!

Lastly, a special thanks to my friends, girlfriend, and family who provided interesting input and continuous support, but most of all, provided a space to clear my mind when it was necessary.

I hope that through this work, I have contributed, even if it is just a little bit, to enhancing safety in driving. I am looking forward to all the technological advances the future has in store for us, and I am very excited to become a part of this world of innovation and creativity.

Bram Kooijman
Delft, October 2021

Contents

Abstract

Previous research showed that perceived risk is an important psychological determinant of road user behaviour and accident prevalence. However, little knowledge exists about how objective in-scene features affect a driver's perceived risk in interactions with pedestrians. This crowdsourcing study tries to fill this research gap. A total of 1082 participants watched 35 out of a total of 86 dashcam videos featuring interactions with pedestrians extracted from the Pedestrian Intention Estimation (PIE) dataset. The videos contained annotations of pedestrian eye contact, crossing behaviour, GPS location, vehicle speed, and yielding rules. The distance between vehicle and pedestrian was manually added, and object counts (detected number of pedestrians, cyclists and vehicles) and respective sizes were added as an index of visual clutter. In each video, participants were asked to press a key on their keyboard and hold it as long as they felt a situation could become risky, and after each video rate perceived risk using a slider and answer whether the pedestrian had made eye contact. Videos in which the participant observed eye contact, increased perceived risk, suggesting that eye contact increases drivers' vigilance. Videos with more visual clutter, and with higher vehicle speed were also associated with increased perceived risk. However, the causality of the correlation with vehicle speed can be questioned and may be mediated by the environment and whether crossing occurred. Videos in which yielding rules were absent, compared to videos in which they were present, did not affect perceived risk. This study is the first to investigate how pedestrians' eye contact affects drivers' perceived risk. The presented results could be useful in safe road design or be used as input for eHMI activation to enhance safety.

Introduction

Yearly about 1.3 million people die as a result of road traffic accidents. 54% of these fatalities involve vulnerable road users (VRUs) (World Health Organisation [WHO], 2018), defined by the World Health Organisation as: "non-motorised road users, such as pedestrians and cyclists as well as motor-cyclists and persons with disabilities or reduced mobility and orientation" (European Commission, 2020). Human error is thought to be the main determinant in between 94% and 96% of traffic crashes (National Highway Traffic Safety Administration [NHTSA], 2017), emphasizing the importance of safety-enhancing measures in traffic (Treat et al., 1979).

Risk perception and road traffic accidents

Deery defined risk perception as "the subjective experience of risk in potential traffic hazards" (Deery, 1999). In the context of road traffic, risk perception is thought to be an important psychological variable as it could predict the behaviour of traffic participants. Moreover, a driver's distorted level of perceived risk is considered one of the main causes of road traffic accidents (Eboli et al., 2017). The psychology behind the effect of perceived risk on driving behaviour can be found in Wilde's Risk Homeostasis Theory, which states that humans compare the amount of risk they perceive to their target level of risk. Consequently, the human compensates in an attempt to eliminate discrepancies between the two (Wilde, 1998). When driving, this discrepancy may cause risky driving behaviour and road accidents (Harbeck & Glendon, 2018; Kouabenan, 2002; Ulleberg & Rundmo, 2003). For example, when drivers perceive lower levels of risk, a common response would be to drive less cautiously (Deery, 1999), e.g., by driving faster (Fuller, 2005). Excessive speed is regarded as highly risky and as an important causation factor in road accidents (Glendon, 2007). An experiment by Renge showed the converse relation that Deery found: when drivers perceive higher risk levels, a lower driving speed is chosen (Renge, 1998).

Risk perception and the characteristics of individuals

Individual differences, such as age and driving experience, are considered to be dominant factors affecting risk perception during driving (Borowsky et al., 2012). For example, more experienced drivers adapt their scanning strategy better according to the traffic environment and show more efficient scanning patterns (Chapman & Underwood, 1998), as compared to less experienced drivers. Also, in a study by Bazilinskyy et al. (2020), cross-cultural differences influenced risk perception due to the participants becoming desensitised to traffic risks from their country of residence.

The effect of visual clutter in road traffic on risk perception

A variety of experiments have been conducted to investigate which detectable on-road features are involved in risk perception. Studies found that increased environmental complexity through certain road design characteristics (i.e., road curvature, lane and shoulder width, gradient, and the presence of median barriers) (Charlton et al., 2014) or dynamic changes involving pedestrians yielded an increase in perceived risk (Cox et al., 2017). Edquist (2009) defined the environmental complexity as 'visual clutter' and identified three different types of visual clutter: *situational clutter* (moving objects on and next to the road that must be attended for safe driving), *designed clutter* (road markings, traffic signs, and signals), and *built clutter* (infrastructure; buildings, billboards, shop signage). It was suggested that high degrees of clutter may impair visual selection, decreasing the ability to detect in-scene changes, and distract the driver by temporarily capturing their attentional resources (Theeuwes et al., 1998). Urban environments are usually characterised by high situational clutter, are therefore inherently perceived as riskier, and also have higher collision frequencies. (Lyon & Persaud, 2002; Miranda-Moreno et al., 2011). Wang et al. (2002) created a probit model (i.e., regression analysis in which the dependent variable can only have two values as outcome) to identify influences on risk perception and found that visual clutter through obstructed visibility and the presence of a potentially conflicting pedestrian were perceived as the riskiest factors in the model.

The effect of eye contact in pedestrian-vehicle interaction on risk perception

Communication during pedestrian-vehicle interaction (PVI) is thought to be an important factor for road safety. Specifically, eye contact with or looking at other traffic participants is thought to help resolve conflicts in PVI (Tartaglia et al., 2019) and serve various functions in traffic. The pedestrian is found to use eye contact as a means to verify mutual awareness (Dey & Terken, 2017), negotiating traffic situations (Möller et al., 2016), or looking for an appropriate gap to cross the road (Yannis, Papadimitriou & Theofilatos, 2013), whereas the driver may use eye contact as a means to estimate pedestrian intention. Recent naturalistic driving studies suggest that when pedestrians communicate with drivers using eye contact, assertive behaviour, and facial expressions, crashes can be prevented (Kong et al., 2021). Rasouli et al. (2018) investigated PVI from the perspective of the driver, by utilizing their self-created dataset, which mainly consisted of crossing situations (with some non-crossing situations) in various geographic locations (including Canada, USA, Germany and Ukraine). The author observed that out of all instances involving pedestrians head movement towards approaching vehicles, 80% occurred before crossing and were therefore marked as a strong indicator of crossing intention. Furthermore, only few studies have been conducted relating looking behaviour of traffic participants to perceived risk. Recent studies indicated that eye contact with the driver from the pedestrian’s perspective can enhance the pedestrian’s feeling of safety (Habibovic et al., 2018; Yang, 2017). A recent crowdsourcing study by Onkhar et al. (2021) showed that eye contact with the driver from the pedestrian’s perspective made people feel considerably safer to cross and that the initiation and termination of eye contact with the driver strongly affected the perceived safety of the pedestrian. In conclusion, eye contact seems to be a promising factor in assessing perceived risk in PVI. However, to the best of our knowledge, no research exists on the effect of pedestrian eye contact on a driver’s perceived risk.

Methods of identifying dynamic influences on perceived risk

Few studies have quantified the driver’s perceived risk as a function of dynamic in-scene variables. Ping et al. created a deep learning model to estimate perceived risk as a function of driving conditions and driving events (Ping et al., 2018) (e.g., obstacles, traffic participant encounters, road characteristics, and vehicle dynamics). A group of 35 participants was asked to assign risk level scores (1–5) to each driving video frame, which were used as input for the deep learning model, together with dynamic and static features extracted from the driving environment. The results showed that the proposed method can effectively model the subjective risk perception behavior of drivers.

Cox et al. investigated the effect of the environment on risk perception using a hazard rating task. 126 participants were presented with 100 pairs of side-by-side images of various traffic situations, which differed in one detail (Cox et al., 2017). On an 11-point scale, the participants had to indicate the riskiness of each change in the images, with examples of in-scene changes being the addition of extra vehicles or traffic signs. Changes involving pedestrians were rated significantly the most dangerous. In rural scenes, changes involving animals were also rated hazardous. Hazard rating were higher when changes involved urban scenes rather than rural scenes.

Wang et al. developed an empirical approach to measure a driver’s perceived safety at roundabouts, using road and traffic characteristics (Wang et al., 2002). 198 participants took part in an oral questionnaire. The attributes of, and traffic at a roundabout were rated on a 5-point Likert scale from very safe to very unsafe to identify attributes affecting the perception of safety. Drivers found that the presence of a potentially conflicting pedestrian, obstructed visibility and increased speed negatively affected perceived safety.

Another method used to measure perceived risk is having participants press a button whenever risk is perceived while watching traffic videos. This method was used before to show that culture (Bazilinskyy, Eisma, Dodou & De Winter, 2020) and driving experience (Borowsky et al., 2012; Scialfa et al., 2011; Wetton et al., 2011) significantly affected perceived risk. Onkhar et al. used this method in combination with crowdsourcing to investigate the effect of eye contact with the driver from the pedestrian’s perspective on the pedestrian’s perceived risk (Onkhar et al., 2021), and found that eye contact made the pedestrian feel considerably safer.

Study aim

This study aims to identify variables, which influence a driver's risk perception in PVI by correlating in-scene variables to measures of perceived risk. The ability of in-vehicle sensors to detect and process dynamic variables in real-time enables the creation of feedback systems to enhance the safety of road users. Therefore, this thesis will focus on real-time measurable in-scene features such as distance to the pedestrian, vehicle speed, object data, yielding rules involved, and eye contact with the pedestrian. The premise of this research is that the identification of factors affecting perceived risk can be of value in detecting and preventing hazardous situations between pedestrians and vehicles. Two measures of perceived risk were collected through a crowdsourcing survey. Participants had to watch a subset of 35 out of a total of 86 videos containing interactions with pedestrians from a vehicle's dashcam. During these videos, the participants were tasked with pressing a response button to indicate risky behaviour and rate the level of risk on a post-trial risk slider. The presented method allows for the real-time online assessment of perceived risk in PVI, using a large set of observable in-scene features, and enabling data collection with a large sample size and spread in demographic characteristics, making it unique for the purpose of this study. Also, previous research indicated that data acquired through crowdsourcing holds equal reliability and validity compared to other approaches (Behrend et al., 2011). The results obtained can reinforce existing knowledge on why and when humans perceive risk in PVI.

Method

Pedestrian Intention Estimation dataset

For the survey, 86 videos were created containing interactions with pedestrians. These videos were extracted from the Pedestrian Intention Estimation (PIE) dataset, a dataset by Rasouli et al. consisting of 10 hours of continuous driving footage through Toronto, Canada (Rasouli et al., 2019), segmented in 60 videos with a duration of 10 minutes. This footage is accompanied by annotations of behavioural variables of the pedestrian (whether the pedestrian is looking towards approaching vehicles and when they start crossing the road), environmental variables (which yielding rules are involved), and vehicle variables (speed of the vehicle and GPS location) with their respective frame of occurrence. All data were recorded from an in-vehicle dashcam installed 1.27 m above the ground with a pitch angle of -10°, a resolution of 1920x1080 px, and a framerate of 30/s.

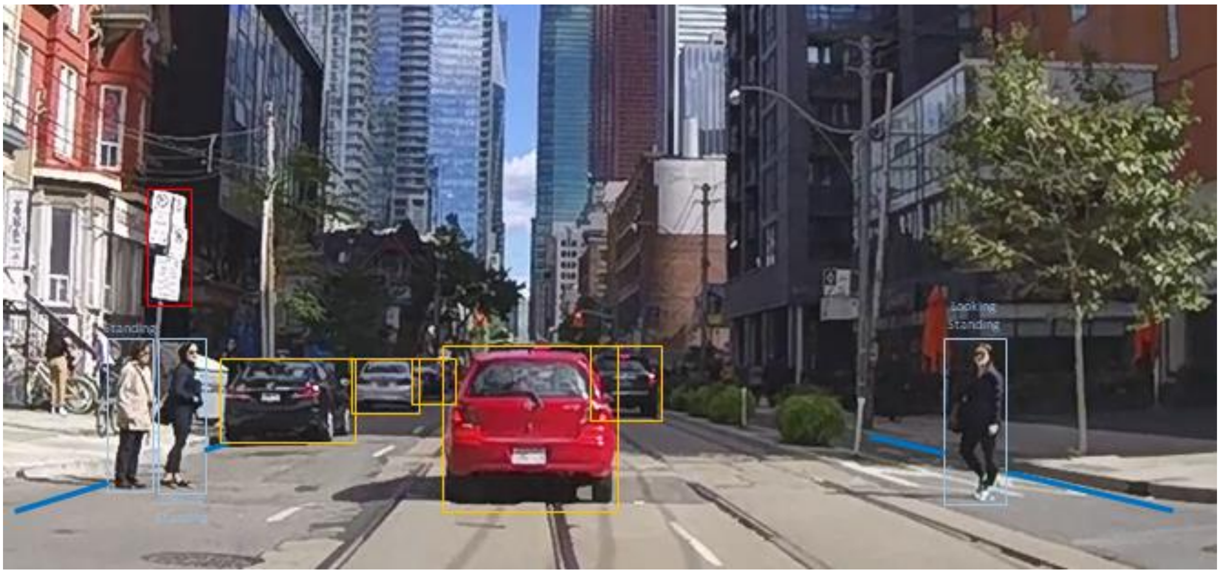

Figure 1. Screenshot of vehicle-pedestrian interaction from the PIE dataset (Rasouli et al., 2019). Bounding boxes depicted in this picture are available in the dataset through coordinates of the most top-left and bottom-right corners of each object.

Video selection

Various interactions with pedestrians were selected from the footage of the PIE dataset, to be extracted as 15 s videos. These videos were normalised around the moment of interaction, which was defined as one of two observable and annotated events expected to play a key role in the perception of risk: a crossing event, initiated when the pedestrian set foot on the road, or an eye contact event, initiated on the first frame the pedestrian of interest looked in the vehicle's direction. Whether one of these events occurred will be referred to henceforth as 'objective crossing' and 'objective looking', respectively. If both objective crossing and objective looking took place, the moment of interaction was defined by the first occurring event. Videos containing diverse sets of parameters were selected to facilitate generalisations of the results for various traffic situations.

Video extraction

The videos were extracted through semi-automated selection. First, python scripts (see Supplementary material) were used to examine the PIE dataset annotations for matches with one of the two key events, which returned the frame on which these events occurred. When matches were found, the video was investigated to verify these annotations. Secondly, a manual inspection through the PIE dataset footage was done to identify additional pedestrian encounters, to include non-annotated events from the dataset. If an encounter without annotations was observed, annotations were manually added through frame-by-frame analysis, using the initiation events mentioned in the previous paragraph. This method was also used to create eye contact and crossing annotations by the creators of the PIE dataset.

After the allocation of events, 15-s videos were extracted from the footage of the PIE dataset. Additionally, a 1-s black screen was added to the beginning of the videos to prevent an abrupt start of the video during the experiment. The resulting videos had a duration of 16 s, with the key event occurring at the 9th second of each video. Rasouli found that the average looking duration was 1.32 s for adults (Rasouli et al., 2018), therefore leaving a 7 s margin after the initiation of the event was expected to be sufficient for the participants to investigate all interactions between vehicle and pedestrian in their entirety. The 8-s pre-event period allows for distinguishing between the interaction phase and a less risky lead-up phase. A total of 86 videos were extracted for this crowdsourcing experiment. Table 1 visualises the combinations of annotations contained in the selected videos.

Table 1.
Visualisation of all combinations of observable in-scene events in the selected videos.

<table>
<tr><td rowspan="14">Videos
(N = 86)</td><td rowspan="8">Crossing
(n = 55)</td><td rowspan="4">Looking
(n = 44)</td><td>No yielding rules (n = 17)</td><td>–</td></tr>
<tr><td rowspan="3">Yielding rules (n = 27)</td><td>Pedestrian crossings (n = 7)</td></tr>
<tr><td>Stop signs (n = 11)</td></tr>
<tr><td>Signalised crossings (n = 9)</td></tr>
<tr><td rowspan="4">Not looking
(n = 11)</td><td>No yielding rules (n = 4)</td><td>–</td></tr>
<tr><td rowspan="3">Yielding rules (n = 7)</td><td>Pedestrian crossings (n = 2)</td></tr>
<tr><td>Stop signs (n = 4)</td></tr>
<tr><td>Signalised crossings (n = 1)</td></tr>
<tr><td rowspan="6">Not crossing
(n = 31)</td><td rowspan="4">Looking
(n = 26)</td><td>No yielding rules (n = 20)</td><td>–</td></tr>
<tr><td rowspan="3">Yielding rules (n = 6)</td><td>Pedestrian crossing (n = 1)</td></tr>
<tr><td>Stop signs (n = 1)</td></tr>
<tr><td>Signalised crossings (n = 4)</td></tr>
<tr><td rowspan="2">Not looking
(n = 5)</td><td>No yielding rules (n = 3)</td><td>–</td></tr>
<tr><td>Yielding rules (n = 2)</td><td>Signalised crossings (n = 2)</td></tr>
</table>

Crowdsourcing experiment

The experiment was conducted using the crowdsourcing platform Appen (www.appen.com). A payment of USD 0.45 was offered for the completion of the experiment. Participants could find the experiment via website channels (e.g., https://www.ysense.com) enlisted among other crowdsourcing jobs available. At the beginning of the experiment, the researchers' contact information was provided. The participants were informed that the purpose of the experiment was "*to determine the risk in situations with pedestrians crossing*". The research was approved by the Human Research Ethics Committee of Delft University of Technology.

First, participants completed a questionnaire containing questions on their demographic characteristics and on-road behaviour (Found in Appendix C). Additionally, one question about communication was asked: "How do you feel about the following?: Communication is important for road safety", where the participants were asked to answer using a 5-point Likert scale from "Completely disagree" to "Completely agree". Lastly, two questions on eye contact were asked, namely "As a driver, what does it mean when a pedestrian makes eye contact with you?" and "As a pedestrian, what does it mean when a driver makes eye contact with you?" with the response option "I should stop", "The pedestrian should stop", "Both should stop", "Neither should stop" and "I prefer not to respond".

After completing the pre-experiment questionnaire, the participants were asked to click a link in Appen to open the webpage with the videos containing the following instruction:

> *You will watch 35 videos of traffic situations involving pedestrians. Some participants will cross the road and some will make eye contact. All videos are recorded within urban Toronto.*
>
> *Each video starts with a black screen. As soon as you see the black screen, press 'F' to start the video.*
>
> *When you feel the situation could become risky, PRESS and HOLD 'F' until you feel the situation is safe again. Press 'F' for any type of risk, including very small risk. You can press and release the key as many times as you want per video. There is no audio involved.*
>
> *The window of your browser should be at least 1300 px wide and 800 px tall.*

Subsequently, a randomised subset of 35 out of the 87 videos was shown to the participant. The resolution of the videos was manually downgraded to 1280x720 px for improved memory allocation while retaining good video quality. After ten trials, a small break was offered, and they could continue to the next batch of videos by pressing the C-key.

After each trial, the participants were asked to respond to the following statement regarding their perceived risk in the previously observed video: "I found the behaviour of the pedestrian(s) to be risky" using a continuous slider from 0 (completely disagree) to 100 (completely agree), as well as to the following statement on subjective eye contact: "The pedestrian(s) made eye contact with me", with response options "Yes", "No" , "Yes but too late", and "I don't know". When both questions were answered, the participants could continue to the next video.

After the last video and associated questions, participants were asked the meaning of four different English-Canadian traffic signs to test basic and English-Canadian traffic sign knowledge. Such knowledge could have played a contextual role in the perception of the situation in the videos observed. Next, the participants were asked to fill in a post-experiment questionnaire in which they would answer the question "Which behaviour increases the feeling of safety?" for the following statements: "Hand gestures increase the feeling of safety", "Early braking increases the feeling of safety", "Eye contact with the pedestrian increases the feeling of safety", "Eye contact with the driver increases the feeling of safety", using four continuous sliders, ranging from 0 (completely disagree) to 100 (completely agree). See Appendix D for instructions, images and questions contained in the survey.

At the end of the experiment, participants were shown a unique code and were asked to note down the code and return to the Appen website. By entering this code in the questionnaire, proof of participation was given, and the payment could be processed.

Additional input variables

By examining the annotated GPS coordinates of the vehicle and comparing them with the vehicle's location in the extracted videos, we found that the available coordinates were not accurate enough to use for data processing. Instead, we used Google Earth to indicate the pedestrian's positions and the vehicle's location at the 9^{th}, 10^{th}, and 11^{th} s of each video (see Appendix B). The moment of interaction occurs at the 9^{th} s of each video, which is why this moment was selected, with the data on the 10^{th} and 11^{th} s providing additional follow-up information. Distances from the vehicle's to the pedestrian's coordinates were approximated using the geodesic distance.

As a parametrisation of visual clutter, Google Cloud Video Intelligence API (Google, 2017) was utilised to detect the number of objects and their respective screen surfaces every 100 ms of each video, where 'objects' refers to all pedestrians, cyclists, and vehicles present in the scene. For the analysis, an additional distinction was made through pedestrians count, and vehicles count. An example of the acquired output, with one of the extracted videos as input, can be found in Appendix B4.

Data analysis and variables

Two measures of risk perception were captured in this study: the participants' keypresses and the post-trial risk slider. Every 100 ms of each video, the mean percentage of keypresses was calculated by averaging data of all participants that had watched that video. Although the videos were normalised around the moment of interaction with the pedestrian, the timing of perceived risk peaks may vary due to the variation in combinations of variables in each video. For example, because driving was frequently done in densely populated areas, multiple interactions with pedestrians might occur in a video. Such variation in environments, encounters, and differences in key events can result in non-uniform distributions of perceived risk throughout the videos, meaning risky events could occur at different moments in time, also before the main moment of interaction. Figure 2 shows an example of two different moments of interaction in the extracted videos. For the analysis, the mean of the total percentage of keypresses in a single video (15 s, excluding the 1 s black screen) was used as a consequence of the differences between videos, because this metric depicts the situations in their entirety rather than situation-specific.

To investigate how eye contact affects perceived risk, a variety of parameters were added to the analyses. First, for each of the post-trial eye contact answers ("Yes", "No", "Yes but too late", "I don't know"), the percentage of participants who had chosen a specific answer was calculated for each video, which allows for investigating subjective eye contact on top of objective eye contact. Second, the percentage of participants who wrongly perceived pedestrian looking behaviour (parameter is called: incorrect looking indications) was calculated to examine the effect of ambiguities and misperceptions of subjective eye contact on perceived risk. Incorrect looking indications were defined as false positives or false negatives in subjective looking behaviour, dependent on the objective data. False positives were calculated by dividing the number of participants who answered "Yes" and "Yes, but too late" by the total number of video responses without objective looking. A similar approach was taken for the false negatives, calculated by dividing the number of participants who responded "No" on a video with the total number of responses to the eye contact question for videos containing objective looking.

Figure 2. An example of two different moments of interactions occurring at the 9-second mark in two of the extracted videos. The left figure takes place at a signalised crossing, with the pedestrian of interest at the right curb, looking at the vehicle and waiting for the light to become green to cross. The right figure shows an interaction with a pedestrian about to cross at a pedestrian crossing, who just turned her head to look at the incoming vehicle.

A Pearson's correlation matrix and linear regression with ordinary least squares were used to estimate the relation between in-scene variables and the two measures of risk perception: the mean number of response keypresses and the mean risk slider values per video. For the in-scene variables that were a function of time (e.g., data stored in 100 ms bins), mean values were calculated per video to be used in the correlation and regression analyses. These time-based variables include distance to the pedestrian, the vehicle speed, number of objects, number of pedestrians, number of vehicles, and object screen surface. The mean values were used because they depict the situations in their entirety rather than at a specific moment. Using moment-specific metrics like maximum or minimum values could result in capturing information that does not correspond to the interaction due to the differences between encounters. Other in-scene variables included in the regression model were the percentage of participants choosing each subjective eye contact answer per video, incorrect looking indications, objective looking and objective crossing, and which vehicle yielding rules were involved (e.g., signs or lights that indicate mandatory vehicle yielding). A summary of all variables involved in the regression and correlation analyses can be found in Table 2. Results were declared significant with a correlation coefficient of $p < 0.05$.

Lastly, Welch's t-tests were conducted between the mean values of the videos' risk-perception measures (keypresses and risk slider) with different yielding rules involved to investigate statistical differences in mean perceived risk values between yielding rules.

Table 2.
Summary of the variables involved in the regression and correlation analysis. All of the variables presented were calculated for each video (N = 86).

	Variable	**Obtained from**
Dynamic variables	Mean distance to the pedestrian* (m)	GPS coordinates obtained from Google Earth
	Mean speed of the vehicle** (km/h)	PIE dataset annotations
	Objective looking (yes/no)	PIE dataset annotations
	Objective crossing (yes/no)	PIE dataset annotations
Visual clutter	Mean object count** (*n*)	Google Cloud Video Intelligence API
	Mean pedestrian count** (*n*)	Google Cloud Video Intelligence API
	Mean vehicle count ** (*n*)	Google Cloud Video Intelligence API
	Mean object surface** (*n*)	Google Cloud Video Intelligence API
Environmental variables	Yielding rules (yes/no)	PIE dataset annotations
	Stop signs (yes/no)	PIE dataset annotations
	Pedestrian crossings (yes/no)	PIE dataset annotations
	Signalised crossings (yes/no)	PIE dataset annotations
	City road (yes/no)	Annotated by the author of this thesis
	Urban (yes/no)	Annotated by the author of this thesis
	Suburban (yes/no)	Annotated by the author of this thesis
Measures of risk perception	Mean risk slider values (0-100)	Obtained through crowdsourcing (post-trial survey)
	Mean of participant's keypresses*** (%)	Obtained through crowdsourcing (in-trial keypresses)
Eye contact	Percentage of participants indicating: *"Yes"* (%)	Obtained through crowdsourcing (post-trial survey)
	Percentage of participants indicating: *"No"* (%)	Obtained through crowdsourcing (post-trial survey)
	Percentage of participants indicating: *"Yes, but too late"* (%)	Obtained through crowdsourcing (post-trial survey)
	Percentage of participants indicating: *"I don't know"* (%)	Obtained through crowdsourcing (post-trial survey)
	Incorrect looking indications (%)	Obtained through crowdsourcing (post-trial survey)

Note. * indicates a mean value calculated using three values; estimations of the distance to the pedestrian on the 9th, 10th and 11th second of each video. ** indicates a mean value that was calculated from 150 values, representing data captured every 100 ms. *** indicates the average percentage of all keypress data captured in a single video.

Results

Data filtering

2361 participants took part in this crowdsourced study between March 17 and June 2, 2021. If the worker ID was not present in both the survey results and the keypress data or was used multiple times, the corresponding participant was excluded (n = 113). Participants, who reported they did not read the instructions (n = 23), who reported being younger than 18 years (n = 3), whose data from less than 35 trials were available, or who had too many videos of unexpected lengths (i.e., less than 25% of data with the length between 15 and 17 s due to connection issues) were removed (n = 123), to counter invalid keypress data. If the survey was executed multiple times from the same IP address, all but the first attempt were removed (n = 570). Lastly, participants who made more than two mistakes in the four questions about Canadian traffic signs (See Appendix D) were also excluded (n = 95). After filtering, data of 1082 participants remained for analysis.

Participants' characteristics

The remaining 1082 participants had a mean age of 34.53 years (*SD* = 10,74), of which 684 were male, 390 were female, and 8 participants preferred not to indicate their gender. The respondents were located in 65 different countries. The five most represented countries were Venezuela (n = 617), USA (n = 117), India (n = 45), Russia (n = 42) and Egypt (n = 35). The mean time of completion of the study was 42.03 min (*SD* = 20.61 min). The average satisfaction score of this crowdsource study was 4.1 on a scale from 1 to 5, based on 69 participants who completed the optional satisfaction survey offered by Appen after the experiment. 702 participants indicated that they had been driving a vehicle at least once a week. 1037 joined the survey from laptops or desktops. The survey was conducted indoors in bright light by 544 participants and indoors in dim light by 363 participants. Private vehicles were used as the main mode of transport for 564 participants. The mean age participants obtained their first drivers' license was 20.89 (*SD* = 5.26). 200 participants indicated that they had driven 0 km in the last 12 months, 241 participants indicated they drove 1–1000 km, whereas 168 and 165 participants, respectively, drove 1000–5000 km and 5000–10000 km.

Participants' opinions on communication in traffic

When the participants were asked about their thoughts about the importance of communication in traffic, 528 participants indicated "*Completely agree*" whereas 336 indicated "*Agree*". When the participants were asked which forms of communication are important for road safety, results were as follows: 88/100 for hand gestures, 86/100 for eye contact with the pedestrian, 85/100 for light signalling to the pedestrian and 84/100 for eye contact with the driver.

The respondents also provided their opinion on the meaning of eye contact between pedestrian and driver. Eye contact with the pedestrian (from the perspective of the driver) and eye contact with the driver (from the perspective of the pedestrian). 468 participants indicated that eye contact with the pedestrian means "*I should stop*", and 413 indicated "*Both should stop*". The meaning of eye contact with the driver was interpreted as "*The driver should stop*" by 378 participants. 365 participants interpreted this behaviour as "*Both should stop*", and 272 participants as "*I should stop*".

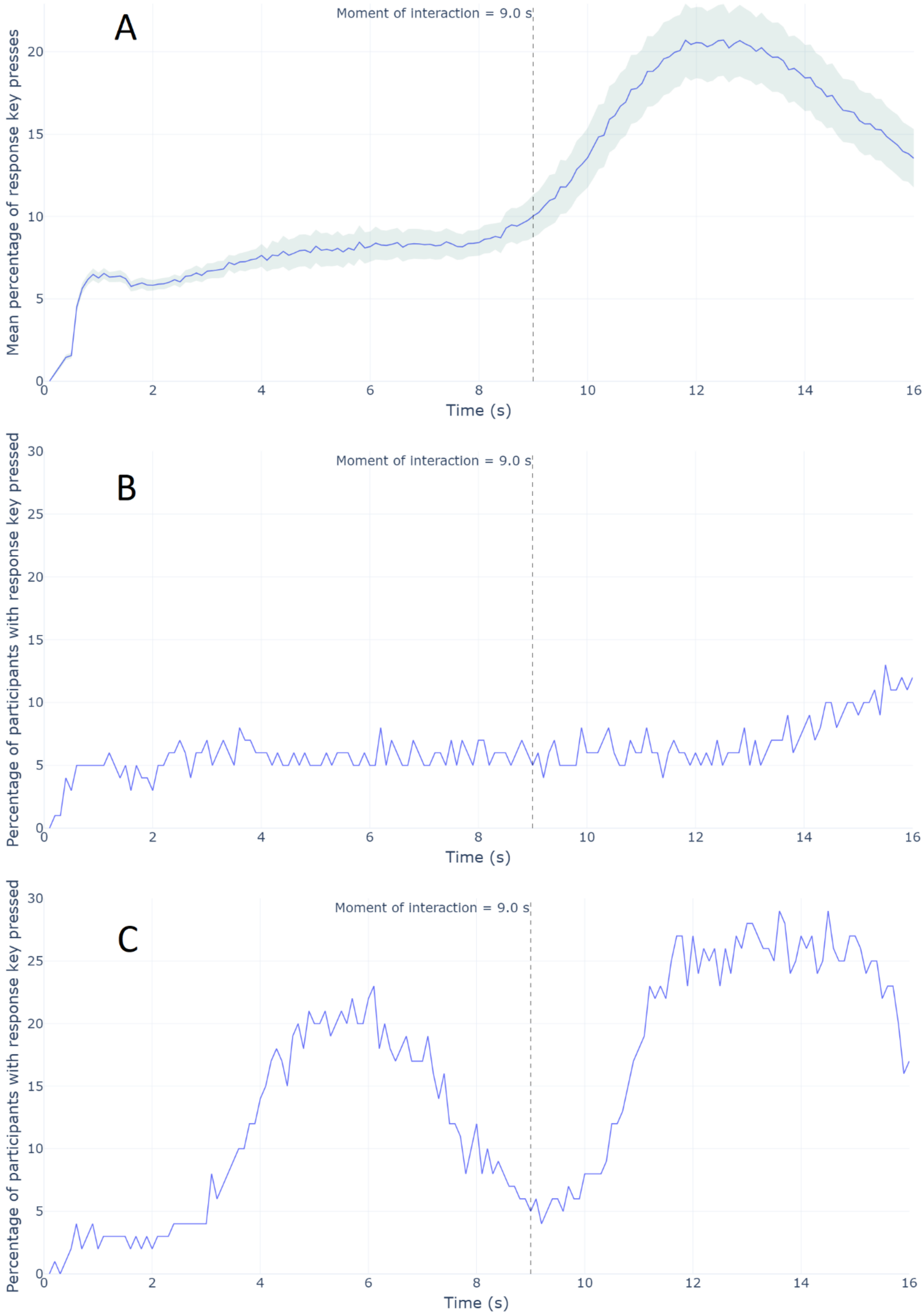

Figure 3. Mean response keypresses over time. (A) The mean keypress response of all videos combined. The error bars represent the standard error of the mean. (B) Response keypresses of an individual video (id = 23): no yielding rules were involved, no objective crossing took place, but objective looking was involved. Overall low risk levels were perceived. (C) Response keypresses of an individual video (id = 1): No yielding rules were present, but both

objective crossing and looking were involved. The peak before the moment of interaction represents an additional interaction with another pedestrian prior to the key event.

Correlation and regression analysis

Figure 3 shows the average keypress responses of all videos combined and the keypress responses of two separate individual videos as a function of time, representing how different sets of variables and events can associate with changes in risk perception. In Figure 4, the Pearson correlation matrix is presented of the input variables and risk measures gathered in the survey, of which Table 3 displays the means and *SD*s.

Various significant correlations with the risk measures were discovered. Correlations found with the mean keypresses included objective crossing (r = 0.57), subjective eye contact: “Yes” (r = 0.35), “No” (r = -0.32), “Yes but too late (r = 0.36)”, “I don’t know” (r = -0.69), the mean vehicle speed (r = -0.52*)*, the mean distance to the pedestrian *(*r = -0.28*)*, the mean object count (r = 0.32), the mean object surface (r = 0.30) the mean pedestrian count (r = 0.39) , pedestrian crossings *(*r = 0.18*)*, and stop signs present (r = 0.24).

Correlations with the mean risk slider were objective looking (r = -0.43), objective crossing (r = 0.57), subjective eye contact: “Yes but too late” (r = 0.42), “I don’t know” (r = -0.65), the mean speed of the vehicle (r = -0.45), the mean pedestrian count (r = 0.26), stop signs present (r = 0.25) and signalised crossings (r = -0.32).

Some other significant correlations that were found in the correlation matrix included “*I don’t know*” with speed (r = 0.55), signalised crossings (r = 0.43), stop signs (r = -0.32) and mean pedestrian count (r = 0.31). Higher object counts were found in urban areas (r = 0.37), and more incorrect looking with less pedestrians (r = 0.19), more distance to the pedestrians (r = 0.22) and at signalised crossings (r = 0.25). The relations with the aforementioned variables and linear regressions will be more elaborately discussed in the following section.

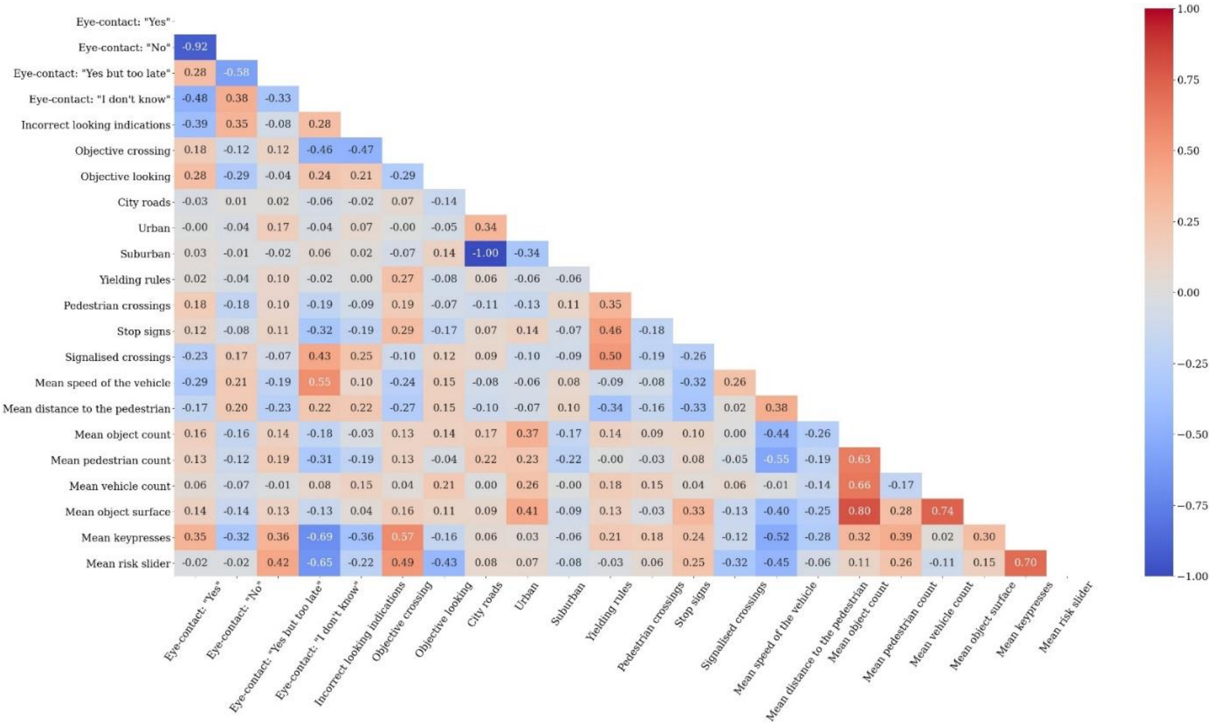

Figure 4. Correlation matrix of the variables included in the experiment. All correlations were determined at the level of videos (N = 86).

Table 3.
The means and standard deviations of all non-categorical variables present in the analysis. All values were calculated at the level of videos (N = 86).

	Mean	***SD***
Mean distance to the pedestrian *(m)*	10.61	6.54
Mean vehicle speed *(km/h)*	15.34	9.32
Mean object count *(n)*	5.62	1.91
Mean pedestrian count *(n)*	2.25	1.44
Mean vehicle count *(n)*	3.30	1.48
Mean object surface *(0-1)*	0.21	0.09
Eye contact: "Yes" *(n)*	136	107
Eye contact: "No" *(n)*	237	121
Eye contact: "I don't know" *(n)*	15	14
Eye contact: "Yes, but too late" *(n)*	44	45
Incorrect looking indications *(%)*	41.20	29.44
Mean risk slider *(0-100)*	47.52	14.50
Mean keypresses *(%)*	12.05	3.78

Tables 4 and 5 show the regression analyses results of the two risk-perception measures with the subjective eye contact data. Additionally, Figure 5 visualises the regression analysis of the mean keypresses with the subjective eye contact data. The answers "Yes but too late" and "I don't know" were both significant predictors in the analysis. When more participants selected these answers, perceived risk increased for "Yes but too late" and decreased for "I don't know". Additionally, when more participants answered "Yes" or "No", perceived risk decreased and increased, respectively, indicating that more evident eye contact increases risk perception. The opposite was true when no clear eye contact was present.

Of all observations of eye contact, on average, 25.26% of the participants had false positive observations of eye contact, whereas 48.74% of the participants provided a false negative observation of eye contact. The correlation matrix shows significant negative correlations of the perceived risk measures with incorrect looking indications, meaning that although plenty of participants frequently misperceived eye contact, the situation was generally perceived as less risky.

Table 4.
Regression results of mean risk slider values versus the frequency of subjective eye contact answers with their respective Pearson correlation coefficient (r) and p-value.

Perceived-risk measure	**Variable**	$\boldsymbol{\beta_0}$	**Coefficient**	**Std err**	***r***	***p***
Mean post-trial risk slider values (0–100)	Yes	46.943	0.004	0.015	0.03	0.774
	No	49.476	-0.008	0.013	0.07	0.530
	I don't know	57.822	-0.675	0.080	-0.65	< **0.001**
	Yes, but too late	41.278	0.141	0.031	0.42	< **0.001**

Note. *N* = 86 videos were used for this regression.

Table 5.
Regression results of mean keypresses versus the frequency of subjective eye contact answers with their respective Pearson correlation coefficient (r) and p-value.

Perceived-risk measure	**Variable**	$\boldsymbol{\beta_0}$	**Coefficient**	**Std err**	***r***	***p***
Mean keypresses (%)	Yes	10.211	0.013	0.004	0.35	< **0.001**
	No	14.627	-0.011	0.003	-0.32	**0.001**
	I don't know	14.831	-0.182	0.020	-0.69	< **0.001**
	Yes, but too late	10.652	0.032	0.008	0.36	< **0.001**

Note. N = 86 videos were used for this regression.

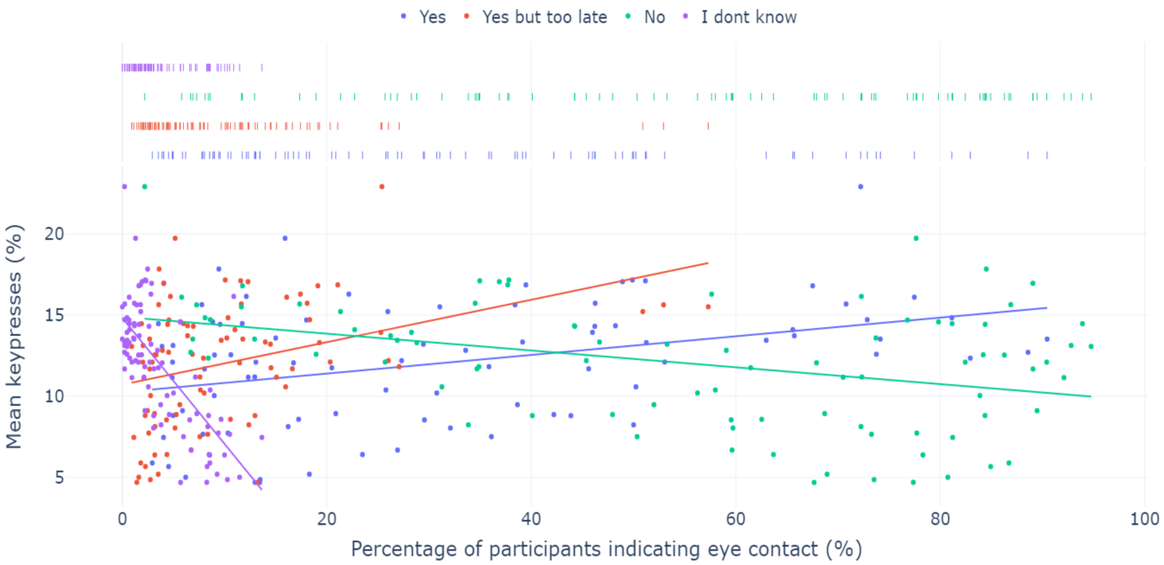

Figure 5. Regression analysis of the mean participants' keypresses versus the percentage of participants indicating each post-trial eye contact answer. Each dot represents one video, with the mean keypresses of all participants per variable on the y-axis, and the percentage of participants indicating each eye contact answer on the x-axis. The blue line represents the answer "Yes", the red line "Yes but too late", the green line "No" and the purple line "I don't know". The following correlations were found: "Yes" (r = 0.35), "Yes, but too late" (r = 0.36), "No" (r = -0.32), and "I don't know" (r = -0.69). N = 86 videos were used for this regression.

Tables 6 and 7 summarise the regression results of the perceived-risk measures with the mean vehicle speed and the mean distance to the pedestrian for each video. Additionally, Figure 6 visualises the regression analysis of the mean keypresses with vehicle speed. Speed was a significant predictor of both perceived-risk measures, indicating that videos with lower vehicle speed yielded higher values of risk perception. Mean distance data were a predictor of mean keypresses only, indicating that the perceived risk increased when the average distance to the pedestrian was smaller.

Table 6.
Regression results of mean risk slider values versus mean vehicle speed and mean distance to the pedestrian with their respective Pearson correlation coefficient (r) and p-value.

Perceived-risk measure	**Variable**	β_0	**Coefficient**	**Std err**	r	p
Mean post-trial risk slider values (0–100)	Mean speed	58.713	-0.711	0.150	-0.65	**< 0.001**
	Mean distance	49.610	-0.136	0.241	-0.07	0.575

Note. N = 85 videos were used for this regression.

Table 7**.**
Regression results of mean keypresses versus mean vehicle speed and mean distance to the pedestrian with their respective Pearson correlation coefficient (r) and p-value.

Perceived-risk measure	**Variable**	β_0	**Coefficient**	**Std err**	r	p
Mean keypresses (%)	Mean speed	15.343	-0.209	0.037	-0.52	**< 0.001**
	Mean distance	13.889	-0.158	0.061	-0.26	**0.011**

Note. N = 85 videos were used for this regression.

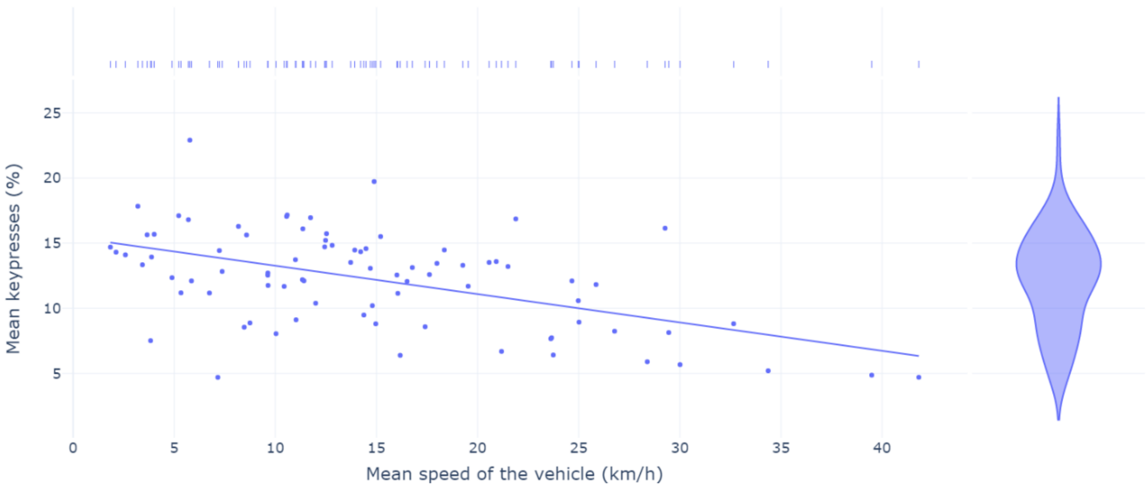

Figure 6. Scatterplot of the mean participants' keypresses versus the mean vehicle speed (r = -0.52). Each dot represents one video (N = 86).

Tables 8 and 9 show the results of the analyses between the mean perceived-risk measures and the mean object data. Additionally, Figure 7 visualises the regression analysis of the mean keypresses with object counts. The mean number of pedestrians present in the scene was a significant predictor of both keypresses and risk slider, indicating that more risk was perceived in videos with more pedestrians. Other significant predictors for the mean keypress data were the total number of objects present and the accumulated size of objects' surfaces, indicating that videos containing more objects and with larger objects' surfaces yielded an increase in perceived risk.

Table 8.
Regression results of mean risk slider values versus visual clutter data with their respective Pearson correlation coefficient (r) and p-value.

Perceived-risk measures	Variable	β_0	Coefficient	Std err	r	p
Mean of post-trial risk slider values (0–100)	Mean object count	41.691	1.037	0.821	0.14	0.210
	Mean pedestrian count	41.579	2.640	1.057	0.26	**0.014**
	Mean vehicle count	50.288	-0.837	1.062	-0.08	0.433
	Mean object surface	41.190	29.025	16.508	0.19	0.082

Note. N = 86 videos were used for this regression.

Table 9.
Regression results of mean keypresses versus visual clutter data with their respective Pearson correlation coefficient (r) and p-value.

Perceived-risk measures	Variable	β_0	Coefficient	Std err	r	p
Mean keypresses (%)	Mean object count	8.277	0.671	0.203	0.32	**0.001**
	Mean pedestrian count	9.704	1.042	0.262	0.39	< **0.001**
	Mean vehicle count	11.77	0.083	0.278	0.03	0.764
	Mean object surface	9.223	12.941	4.151	0.30	**0.002**

Note. N = 86 videos were used for this regression.

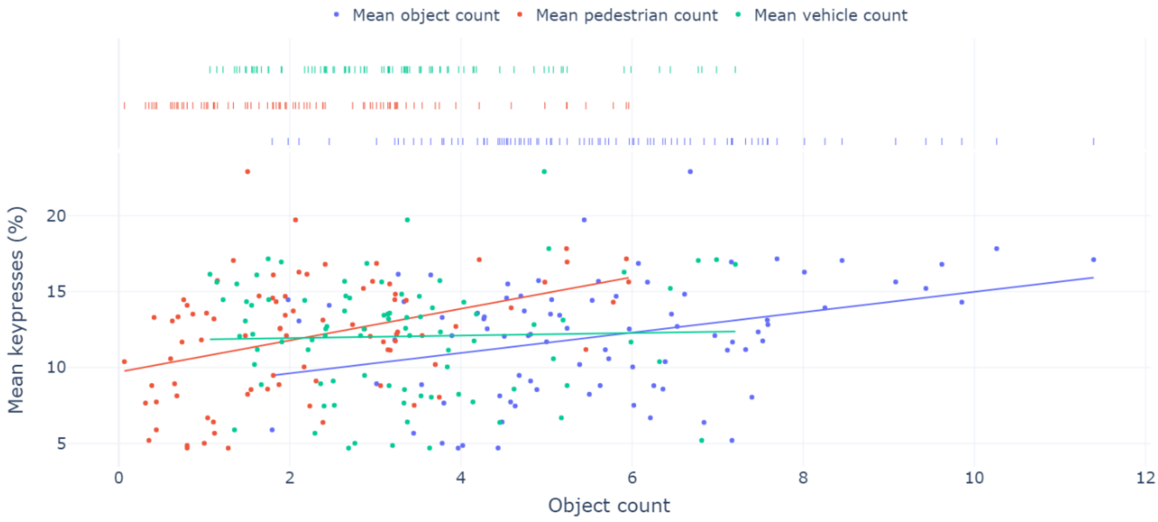

Figure 7. Scatterplot of the mean keypresses vs mean object counts. Each dot represents one video (*N* = 85). The *r*- values for object count, pedestrian count and vehicle count were respectively: *r* = 0.32, *r* = 0.39, *r* = 0.03.

No significant results were found between the perceived-risk measures and videos without yielding rules. To investigate significant differences in risk perception between the videos involving different yielding rules, Welch's *t*-tests were conducted.

The results of Welch's *t*-tests are shown in Tables 10 and 11. For the slider, significant differences were found between videos without yielding rules versus signalised crossings, stop signs versus signalised crossings, and signalised crossings versus pedestrian crossings. For the keypresses, videos without rules versus pedestrian crossings, no yielding rules versus stop signs, and signalised crossings versus pedestrian crossings yielded significant differences.

Table 10.
Results of Welch's t-tests displaying significant differences between mean risk slider values of videos containing different yielding rules, with their respective t and p-value.

Perceived-risk measures	**Rule #1**	**Rule #2**	**df**	***t***	***p***
Mean post-trial risk slider values (0–100)	None	Pedestrian crossings	46.837	-1.033	0.306
	None	Signalised crossings	36.530	2.224	**0.032**
	None	Stop signs	33.332	-1.968	0.058
	Stop signs	Pedestrian crossings	16.897	1.593	0.130
	Stop signs	Signalised crossings	32.990	3.683	**< 0.001**
	Signalised crossings	Pedestrian crossings	19.956	3.406	**0.003**

Note. Videos of yielding rules included pedestrian crossings (*n* = 10), stop signs (*n* =16), signalised crossings (*n* = 19) and no rules involved (*n* = 41).

Table 11.
Results of Welch's t-tests displaying significant differences between mean keypresses of videos containing different yielding rules, with their respective t and p-value.

Perceived-risk measures	**Rule #1**	**Rule #2**	**df**	***t***	***p***
Mean keypresses (%)	None	Pedestrian crossings	20.259	-3.087	**0.006**
	None	Signalised crossings	26.746	0.187	0.853
	None	Stop signs	32.514	-3.049	**0.005**
	Stop signs	Pedestrian crossings	22.456	-0.041	0.968
	Stop signs	Signalised crossings	29.979	2.244	0.032
	Signalised crossings	Pedestrian crossings	26.850	2.273	**0.031**

Note. Videos of yielding rules included pedestrian crossings (*n* = 10), stop signs (*n* =16), signalised crossings (*n* = 19) and no rules involved (*n* = 41).

Discussion

In this crowdsourcing study, participants were asked to view vehicle dashcam videos of interactions with pedestrians in various traffic environments and were tasked with pressing a response button to indicate risky behaviour. Two risk measures were acquired: keypresses of the participants when risk was detected and a post-trial risk slider, which allowed for examining how perceived risk varies as a function of objective detectable in-scene variables. Through correlation and regression analyses, numerous variables significantly affecting perceived risk were identified and quantified.

Through the post-trial eye contact question, data were obtained on the importance of pedestrians' eye contact on drivers' perceived risk. The answer "I don't know" was associated with a decrease in perceived risk, of which a likely explanation is the relative unimportance of the pedestrian in the scene, for example due to the absence of a crossing event or the presence of a signalised crossing. However, the post-trial answer "I don't know" on the instruction "*The pedestrian(s) made eye contact with me*" might have have been interpreted as "*I don't know which pedestrian you mean"*, instead of "*I don't know if the pedestrian made eye-contact*" causing some ambiguity in the meaning of this answer. The answer "Yes but too late" was associated with an increase in perceived risk, emphasising the importance of the timing of looking and head orientation, which are thought to be important signs of pedestrian awareness (Wang et al., 2020). Thompson et al. found that distracted pedestrians are more frequently associated with displaying unsafe behaviour (Thompson et al., 2013). Therefore, late looking as an indicator of inattention might induce the feeling of risk. Lastly, when more participants answered with "Yes" or "No", perceived risk increased and decreased, respectively. Although eye contact with the driver from the pedestrian's perspective was found to decrease perceived risk (Onkhar et al., 2021), this study established the inverse relation from the driver's perspective. Rasouli et al. showed that when pedestrians looked at an incoming vehicle it was often an indication of crossing intentions (Rasouli et al., 2018). The participants may also use the looking behavior of the participant to predict if an interaction will occur. The uncertainty of whether an interaction will occur, and the interaction itself could consequently increase perceived risk. Following Wilde's risk theory (Wilde, 1998), this increase in perceived risk can increase a driver's vigilance, meaning that pedestrian eye contact is important for enhancing safe driving. This argumentation would confirm pedestrian eye contact as an important means of communication with the driver, as indicated by various studies (Kong et al., 2021; Rasouli et al., 2018; Tartaglia et al., 2019).

When comparing objective looking to subjective looking, large discrepancies were established, in which objective looking was often perceived as not looking. Eye contact might have gone unnoticed by the participant because of high visual clutter, the participant might have missed the moment of looking due to looking occurring distant from the vehicle, or the gaze length was very short. Additionally, eye contact could have been missed due to a lack of immersion, affecting the participants' focus and hazard detection capabilities. The negative correlation of objective looking with risk perception and the positive correlation of objective looking with "I don't know" could implicate that misperceived looking more frequently occurred in situations where looking was less important due to the situation being less risky. It could be argued that humans are generally good at filtering irrelevant information (Bazilinskyy, Kooijman, Dodou, De Winter, 2020; Decker et al., 2015; Yannis, Papadimitriou, Papantoniou & Voulgari, 2013), meaning that an irrelevant pedestrian could result in eye contact not being observed, and supports the claim of the situation being less risky.

Overall, more risk was perceived in videos with more visual clutter, specifically situational clutter, through object density and pedestrian density, which is consistent with previous findings (Cox et al., 2017; Edquist, 2009). A possible explanation for these results is that an increase in situational clutter makes it more difficult for drivers to allocate attentional resources efficiently, resulting in information overload (Lerner et al., 2003) and increasing perceived risk.

An unexpected finding was that videos with lower mean vehicle speed were perceived as riskier. A higher vehicle speed was observed in locations with less situational clutter and often in the presence of roads involving signalised crossings, which both correlated to a decrease in perceived risk due to inherently decreasing the probability of an interaction between vehicle and pedestrian. In parallel, the negative correlation between speed and "I don't know" could indicate that the presence of the pedestrian is relatively insignificant at higher speeds. Speed seems to be a function of location (e.g., higher speeds at signalised crossings), objective crossing value (e.g., higher speeds with non-crossing pedestrians), and situational clutter rather than an indicator of perceived risk. A study by Hoogmoed (Hoogmoed, 2021) found a similar correlation of vehicle speed with perceived risk and also suggested environmental features to be the main influence rather than the relation itself being causal. Because speeding is an important causal factor in accident involvement (Baldwin et al., 2018), it might be interesting to distinguish between relative and absolute speed in subsequent research by comparing the vehicle's speed to the allowed speed limit or the speed of surrounding vehicles.

Videos in which smaller distances to the pedestrian were observed yielded an increase in perceived risk. In a crossing situation, a shorter distance to the pedestrian may be equal to an interaction with a pedestrian. Because

PVI itself is thought to increase perceived risk, this result seems logical. However, as a consequence of incorrect vehicle GPS data in the PIE dataset, only limited data on distance were available. Furthermore, the distance data concerned one pedestrian of interest, that is, interactions with other pedestrians were not accounted for. This means that peaks in risk perception could have a mismatch with the corresponding distance profile (as was depicted in the risk profile seen in Figure 3C).

Welch's *t*-tests show various significant differences between the risk scores of yielding rules. One way to interpret these results is by looking at the characteristics associated with these yielding rules. First, the presence of a signalised crossing decreases situational ambiguities, as pedestrians at signalised crossings are presumed not to cross when their light is red (as well as that the driver of the vehicle is expected to stop for a red light), therefore inherently decreasing perceived risk. Although stop signs and pedestrian crossings also enforce strict rules, both of these show an increase in perceived risk. At both these yielding rules, the behaviour of the driver is dependent on the behaviour of the pedestrian. For example, at pedestrian crossings the assertiveness of the pedestrian is an important determinant in negotiating traffic situations in PVI (Kong et al., 2021; Salamati et al., 2013), and therefore miscommunication between a driver and a pedestrian can easily cause a conflict and increase the possibility of ambiguities. Similar reasoning is possible for situations with stop signs involved. Hence, both stop signs and pedestrian crossings could inherently increase perceived risk, because of the increased probability of PVI. Lastly, although the lack of traffic signs and traffic lights have previously been identified as detrimental for safe use of roads (Aceves-González et al., 2020), no significant correlation was found between perceived risk and the absence of yielding rules. A possible explanation for this result is that pedestrians behave more safely, for example, by maintaining more distance or communicating intentions more clearly, to compensate for the lack of yielding rules involved. This explanation would be in line with Dey and Terken (2017), showing that pedestrians have increased awareness on roads without pedestrian crossing due to knowing that they do not have the right of way.

The current paper presents a unique study, by investigating perceived risk in PVI in real-time from the driver's perspective, with a large feature-set detectable by in-vehicle sensors, and large sample size. It is also the first to investigate how pedestrians' eye contact affects drivers in terms of perceived risk. A small group of studies did have a somewhat similar approach. Hoogmoed (2021) also used crowdsourcing to assess perceived risk from the driver's perspective. However, in Hoogmoed's study, participants assessed perceived risk in images rather than the real-time assessment of perceived risk in videos. Onkhar et al. (2021) used the same approach used in the current study, but from the pedestrian's perspective and with the videos being created through a simulator rather than naturalistic driving videos. Also, Onkhar's study focused solely on the effect of eye contact instead of other additional detectable in-scene features. Cabrall et al. (2020) did look at the effect of various visible driving scene features, but examined how these features affect eye parameters (i.e., saccade amplitude, pupil diameter and fixation duration) as a parametrisation of subjective effort rather than perceived risk. Additionally, Compared to Cabrall and most other studies that investigated perceived risk, this study offers a large increase in sample size.

Limitations and recommendations

The use of crowdsourcing introduces some limitations to consider. First, there is a lack of control of the state in which the participant conducts the survey, possibly affecting the genuineness and authenticity of the participants' data. Additionally, the method used may lack immersion, affecting perceptual fidelity and causing doubt in how the obtained results transfer to a real-world driving scenario. Reasons for this doubt are that the participant was not occupied with the task of driving a vehicle, causing the participant not actually to be at risk, thus not having to deal with the consequences of mistakes, and not having to divide attention between both the task of driving and perceiving the environment. How immersive the survey was, was also affected by the hardware that the participants used. Differences in hardware between participants have previously been thought to cause some variation in the participant's data (Bazilinskyy & De Winter, 2018). Using a simulator with other forms of feedback could improve perceptual fidelity, but at the cost of sample size, and variation of demographic make-up of the participant group. The lack of audio could also affect the perceptual fidelity of the survey. Including audio might give interesting complementary results, as it is shown that multimodal feedback can enhance performance (Vitense et al., 2003).

Next, how participants perceive risk is directly affected by their ability to detect hazards. A separation between risk perception and hazard detection skills could be added by combining the current study with a hazard perception test (Scialfa et al., 2011; Wetton et al., 2011) to quantify hazard detection capabilities and relate these to perceived risk. One of the problems using these hazard tests in combination with crowdsourcing is that hazard detection test scores might not be valid when the participant's culture differs from the culture in which the test was developed (Bazilinskyy, Eisma, Dodou & De Winter, 2020). In the current study more than half of the participants were Venezuelan, meaning that most of the results represent a culturally different subset of the world compared to where the actual driving in the videos occurred (Toronto, Canada). So, would the obtained results generalise well for a subset of the world population? Through the question on the meaning of certain traffic signs (see Appendix D), the participants have shown to some extend basic knowledge of traffic rules, English-Canadian signs to be specific, and are therefore expected to at least know what is happening in the scene. Also, previous studies with similar participant samples show that results generalise well between participants from other countries (Bazilinskyy et al., 2021; Onkhar et al., 2021). However, some differences in risk perception due to cultural discrepancies will remain.

Other limitations lie in the used dataset. First, only a limited number of combinations of parameters were available in the PIE dataset as a function of the size of the dataset. Second, annotated vehicle location data was inaccurate. Although location data was manually added, significantly fewer data were available than other variables (3 data points instead of 150 per video) and the manual annotating of vehicle and pedestrian locations may have introduced a small error.

To fully understand how perceived risk affects the interaction between vehicle and pedestrian, both the driver's and the pedestrian's perspectives must be comprehended. Traffic involves multidirectional communication, and therefore both the inputs and outputs of the driver and pedestrian are important in resolving conflict. An interesting idea would be to conduct a similar study from the perspective of the pedestrian. Such a study would enable us to match the driver's perceived risk to the pedestrian's perceived risk, which could offer interesting insights and show discrepancies between risk perception of road users. For example, knowing when an approaching pedestrian perceives high risk could function as an activation threshold for external human-machine interfaces. Unfortunately, such a dataset does not yet exist and therefore would need to be created first.

Conclusion

The present study successfully established various detectable in-scene variables, influencing perceived risk in the interaction between pedestrians and vehicles. Additionally, This study is the first to quantify how pedestrians' eye contact affects drivers' perceived risk. The results found in this study could be useful in safe road design or be used as input for activating an external human-machine interface to enhance safety. The following results were obtained:

- Videos in which eye contact was observed by the participants yielded an increase in perceived risk.
- Videos with higher vehicle speed yielded an increase in perceived risk. However, the causality of this correlation was questioned and may be mediated by visual clutter, road characteristics and yielding rule characteristics.
- Videos with more visual clutter yielded an increase in perceived risk.
- Videos in which yielding rules were absent, compared to videos in which they were present, did not affect perceived risk.
- Videos in which the distance between vehicle and pedestrian was smaller, yielded an increase in perceived risk.

Supplementary material

Python code which was used to create and analyse data can be found online at: https://github.com/bazilinskyy/crossing-crowdsourcing. The code is divided in 3 parts: Video extraction and annotation matching, Object detection, and survey data parsing.

The anonymised data that was used to generate the results in this study can be accessed through the following link: https://www.dropbox.com/sh/l1xuoe2vcf2lyw2/AABuUxLCCZpkQyDYCXB21BeRa?dl=0

References

Aceves-González, C., Ekambaram, K., Rey-Galindo, J., & Rizo-Corona, L. (2020). The role of perceived pedestrian safety on designing safer built environments. *Traffic Injury Prevention*, *21*(S1), S84–S89. https://doi.org/10.1080/15389588.2020.1812062

Baldwin, C. L., Lewis, B. A., Greenwood, P. M., Baldwin, C. L., Lewis, B. A., & Greenwood, P. M. (2018). Pedestrians and cyclists 2018. *European Road Safety Observatory*, 87–102.

Bazilinskyy, P., Eisma, Y. B., Dodou, D., & de Winter, J. C. F. (2020). Risk perception: A study using dashcam videos and participants from different world regions. *Traffic Injury Prevention*, *0*(0), 1–7. https://doi.org/10.1080/15389588.2020.1762871

Bazilinskyy, P., & De Winter, J. C. F. (2018). Crowdsourced Measurement of Reaction Times to Audiovisual

Stimuli With Various Degrees of Asynchrony. *Human Factors*, *60*(8), 1192–1206. https://doi.org/10.1177/0018720818787126
Bazilinskyy, P., Dodou, D., Eisma, Y. B., De Winter, J., Bazilinskyy, P., Dodou, D., Eisma, Y. B., Vlakveld, W., & De Winter, J. C. F. (2021). *Blinded windows and empty driver seats: The effects of automated vehicle characteristics on cyclists' decision-making Blinded windows and empty driver seats: The effects of automated 1 vehicle characteristics on cyclists' decision-making 2 3*. *April*. https://www.researchgate.net/publication/342637884
Bazilinskyy, P., Kooijman, L., Dodou, D., & De Winter, J. C. F. (2020). *How should external Human-Machine Interfaces behave? Examining the effects of colour, position, message, activation distance, vehicle yielding, and visual distraction among 1,434 participants*. *April*. https://doi.org/10.13140/RG.2.2.11345.56164
Behrend, T. S., Sharek, D. J., Meade, A. W., & Wiebe, E. N. (2011). The viability of crowdsourcing for survey research. *Behavior Research Methods*, *43*(3), 800–813. https://doi.org/10.3758/s13428-011-0081-0
Borowsky, A., Oron-Gilad, T., Meir, A., & Parmet, Y. (2012). Drivers' perception of vulnerable road users: A hazard perception approach. *Accident Analysis and Prevention*, *44*(1), 160–166. https://doi.org/10.1016/j.aap.2010.11.029
Cabrall, C. D. D., Happee, R., & de Winter, J. C. F. (2020). Prediction of effort and eye movement measures from driving scene components. *Transportation Research Part F: Traffic Psychology and Behaviour*, *68*, 187–197. https://doi.org/10.1016/j.trf.2019.11.001
Chapman, P. R., & Underwood, G. (1998). Visual search of driving situations: Danger and experience. *Perception*, *27*(8), 951–964. https://doi.org/10.1068/p270951
Charlton, S. G., Starkey, N. J., Perrone, J. A., & Isler, R. B. (2014a). What's the risk? A comparison of actual and perceived driving risk. *Transportation Research Part F: Traffic Psychology and Behaviour*, *25*(PART A), 50–64. https://doi.org/10.1016/j.trf.2014.05.003
Charlton, S. G., Starkey, N. J., Perrone, J. A., & Isler, R. B. (2014b). What's the risk? A comparison of actual and perceived driving risk. *Transportation Research Part F: Traffic Psychology and Behaviour*, *25*(PART A), 50–64. https://doi.org/10.1016/j.trf.2014.05.003
Cox, J. A., Beanland, V., & Filtness, A. J. (2017). Risk and safety perception on urban and rural roads: Effects of environmental features, driver age and risk sensitivity. *Traffic Injury Prevention*, *18*(7), 703–710. https://doi.org/10.1080/15389588.2017.1296956
Decker, J. S., Stannard, S. J., McManus, B., Wittig, S. M. O., Sisiopiku, V. P., & Stavrinos, D. (2015). The Impact of Billboards on Driver Visual Behavior: A Systematic Literature Review. *Traffic Injury Prevention*, *16*(3), 234–239. https://doi.org/10.1080/15389588.2014.936407
Deery, H. A. (1999). Hazard and Risk Perception among Young Novice Drivers. *Journal of Safety Research*, *30*(4), 225–236. https://doi.org/10.1016/S0022-4375(99)00018-3
Dey, D., & Terken, J. (2017). Pedestrian interaction with vehicles: Roles of explicit and implicit communication. *AutomotiveUI 2017 - 9th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications, Proceedings*, 109–113. https://doi.org/10.1145/3122986.3123009
Eboli, L., Mazzulla, G., & Pungillo, G. (2017). How to define the accident risk level of car drivers by combining objective and subjective measures of driving style. *Transportation Research Part F: Traffic Psychology and Behaviour*, *49*, 29–38. https://doi.org/10.1016/j.trf.2017.06.004
Edquist, J. (2009). The Effects of Visual Clutter on Driving Performance. *Systems Man and Cybernetics 1999 IEEE SMC99*, 226. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=823381
European Commission. (2020). *Supporting Study on of the New Working Programme of the Its*.
Fuller, R. (2005). Towards a general theory of driver behaviour. *Accident Analysis and Prevention*, *37*(3), 461–472. https://doi.org/10.1016/j.aap.2004.11.003
Glendon, A. I. (2007). Driving violations observed: An Australian study. *Ergonomics*, *50*(8), 1159–1182. https://doi.org/10.1080/00140130701318624
Habibovic, A., Lundgren, V. M., Andersson, J., Klingegård, M., Lagström, T., Sirkka, A., Fagerlönn, J., Edgren, C., Fredriksson, R., Krupenia, S., Saluäär, D., & Larsson, P. (2018). Communicating intent of automated vehicles to pedestrians. *Frontiers in Psychology*, *9*(AUG). https://doi.org/10.3389/fpsyg.2018.01336
Harbeck, E. L., & Glendon, A. I. (2018). Driver prototypes and behavioral willingness: Young driver risk perception and reported engagement in risky driving. *Journal of Safety Research*, *66*, 195–204. https://doi.org/10.1016/j.jsr.2018.07.009
Hoogmoed, J. (2021). Human driver risk perception model: Fundamental threat parameters and what makes a driving situation risky. available from *https://repository.tudelft.nl/*
Kong J., Das, S., Zhang, Y., & Xiao, X. (2021). Lessons learned from pedestrian-driver communication and yielding patterns. *Transportation Research Part F: Traffic Psychology and Behaviour*, *79*, 35–48. https://doi.org/10.1016/j.trf.2021.03.011
Kouabenan, D. R. (2002). Occupation, driving experience, and risk and accident perception. *Journal of Risk*

Research, *5*(1), 49–68. https://doi.org/10.1080/13669870110036577
Lerner, N. D., Llaneras, R. E., McGee, H. W., Taori, S., & Alexander, G. (2003). *Additional Investigations on Driver Information Overload*. 1–25.
Lyon, C., & Persaud, B. (2002). Pedestrian collision prediction models for urban intersections. *Transportation Research Record*, *d*(1818), 102–107. https://doi.org/10.3141/1818-16
Miranda-Moreno, L. F., Morency, P., & El-Geneidy, A. M. (2011). The link between built environment, pedestrian activity and pedestrian-vehicle collision occurrence at signalized intersections. *Accident Analysis and Prevention*, *43*(5), 1624–1634. https://doi.org/10.1016/j.aap.2011.02.005
Möller, L., Risto, M., & Emmenegger, C. (2016). The social behavior of autonomous vehicles. *UbiComp 2016 Adjunct - Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 686–689. https://doi.org/10.1145/2968219.2968561
NHTSA. (2017). 2016 fatal motor vehicle crashes: Overview. *Traffic Safety Facts*, *October*, 1–9.
Onkhar, V., Bazilinskyy, P., Dodou, D., & Winter, J. De. (2021). *The effect of drivers ' eye contact on pedestrians ' perceived safety The effect of drivers ' eye contact on pedestrians ' perceived safety*. *June*.
Ping, P., Sheng, Y., Qin, W., Miyajima, C., & Takeda, K. (2018). Modeling Driver Risk Perception on City Roads Using Deep Learning. *IEEE Access*, *6*, 68850–68866. https://doi.org/10.1109/ACCESS.2018.2879887
Rasouli, A., Kotseruba, I., Kunic, T., & Tsotsos, J. (2019). PIE: A large-scale dataset and models for pedestrian intention estimation and trajectory prediction. *Proceedings of the IEEE International Conference on Computer Vision*, *2019-Octob*, 6261–6270. https://doi.org/10.1109/ICCV.2019.00636
Rasouli, A., Kotseruba, I., & Tsotsos, J. K. (2018). Understanding Pedestrian Behavior in Complex Traffic Scenes. *IEEE Transactions on Intelligent Vehicles*, *3*(1), 61–70. https://doi.org/10.1109/TIV.2017.2788193
Renge, K. (1998). Drivers' Hazard and Risk Perception, Confidence in Safe Driving, and Choice of Speed. *IATSS Res.*, *22*(2), 103–110.
Salamati, K., Schroeder, B., Geruschat, D., & Rouphail, N. (2013). Event-based modeling of driver yielding behavior to pedestrians at two-lane roundabout approaches. *Transportation Research Record*, *2013*(2389), 1–11. https://doi.org/10.3141/2389-01
Scialfa, C. T., Deschênes, M. C., Ference, J., Boone, J., Horswill, M. S., & Wetton, M. (2011). A hazard perception test for novice drivers. *Accident Analysis and Prevention*, *43*(1), 204–208. https://doi.org/10.1016/j.aap.2010.08.010
Tartaglia, R., Albolino, S., Alexander, T., & Fujita, Y. (2019). 20th Congress of the International Ergonomics Association, IEA 2018. In *Advances in Intelligent Systems and Computing* (Vol. 824, Issue Iea).
Theeuwes, J., Kramer, A. F., Hahn, S., & Irwin, D. E. (1998). Our Eyes Do Not Always Go Where We Want Them to Go: Capture of the Eyes by New Objects. *Psychological Science*, *9*(5), 379–385. https://doi.org/10.1111/1467-9280.00071
Thompson, L. L., Rivara, F. P., Ayyagari, R. C., & Ebel, B. E. (2013). Impact of social and technological distraction on pedestrian crossing behaviour: An observational study. *Injury Prevention*, *19*(4), 232–237. https://doi.org/10.1136/injuryprev-2012-040601
Treat, J. R., Tumba, N. S., McDonald, S. T., Shinar, D., Hume, R. D., Mayer, R. E., Stansifer, R. L., & Castellan, N. J. (1979). Tri-Level Study of the Causes of Traffic Accidents: An overview of final results. *Proc. of the American Association for Automotive Medicine Annual Conference*, *21*, 391–403.
Ulleberg, P., & Rundmo, T. (2003). Personality, attitudes and risk perception as predictors of risky driving behaviour among young drivers. *Safety Science*, *41*(5), 427–443. https://doi.org/10.1016/S0925-7535(01)00077-7
Vitense, H. S., Jacko, J. A., & Emery, V. K. (2003). Multimodal feedback: An assessment of performance and mental workload. *Ergonomics*, *46*(1–3), 68–87. https://doi.org/10.1080/00140130303534
Wang, B., Hensher, D. A., & Ton, T. (2002). Safety in the road environment: A driver behavioural response perspective. *Transportation*, *29*(3), 253–270. https://doi.org/10.1023/A:1015661008598
Wang, H., Zhang, H., Gu, M., & Wang, C. (2020). Comparison of the pedestrian crossing intention parameters and research on intention recognition model under different road conditions. *IEEE Access*, *8*, 196675–196689. https://doi.org/10.1109/ACCESS.2020.3034236
Wetton, M. A., Hill, A., & Horswill, M. S. (2011). The development and validation of a hazard perception test for use in driver licensing. *Accident Analysis and Prevention*, *43*(5), 1759–1770. https://doi.org/10.1016/j.aap.2011.04.007
WHO. (2018). Global Status Report on Road. *World Health Organization*, 20. https://www.who.int/publications/i/item/9789241565684
Wilde, G. J. S. (1998). Risk homeostasis theory: An overview. *Injury Prevention*, *4*(2), 89–91. https://doi.org/10.1136/ip.4.2.89
Yang, S. U. (2017). Driver behavior impact on pedestrians' crossing experience in the conditionally autonomous driving context. *Degree Project Computer Science and Engineering*, 11.

http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-220545
Yannis, G., Papadimitriou, E., & Theofilatos, A. (2013). Pedestrian gap acceptance for mid-block street crossing. *Transportation Planning and Technology*, *36*(5), 450–462. https://doi.org/10.1080/03081060.2013.818274
Yannis, G., Papadimitriou, E., Papantoniou, P., & Voulgari, C. (2013). A statistical analysis of the impact of advertising signs on road safety. *International Journal of Injury Control and Safety Promotion*, *20*(2), 111–120. https://doi.org/10.1080/17457300.2012.686042

Appendix A – Additional analysis

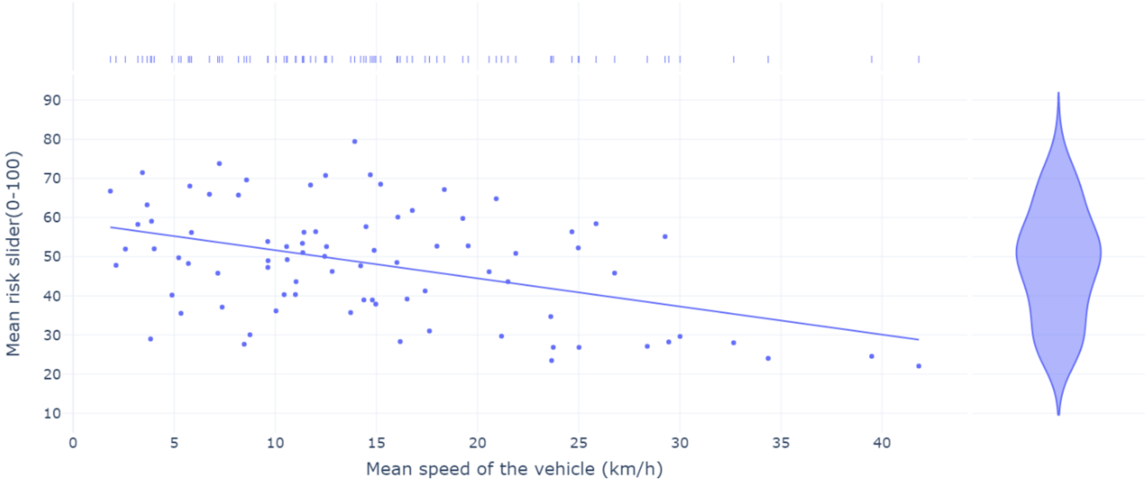

Figure A1. Scatterplot of the post-trial risk slider vs mean vehicle speed ($r = 0.65$). Each dot represents one video ($N = 86$)

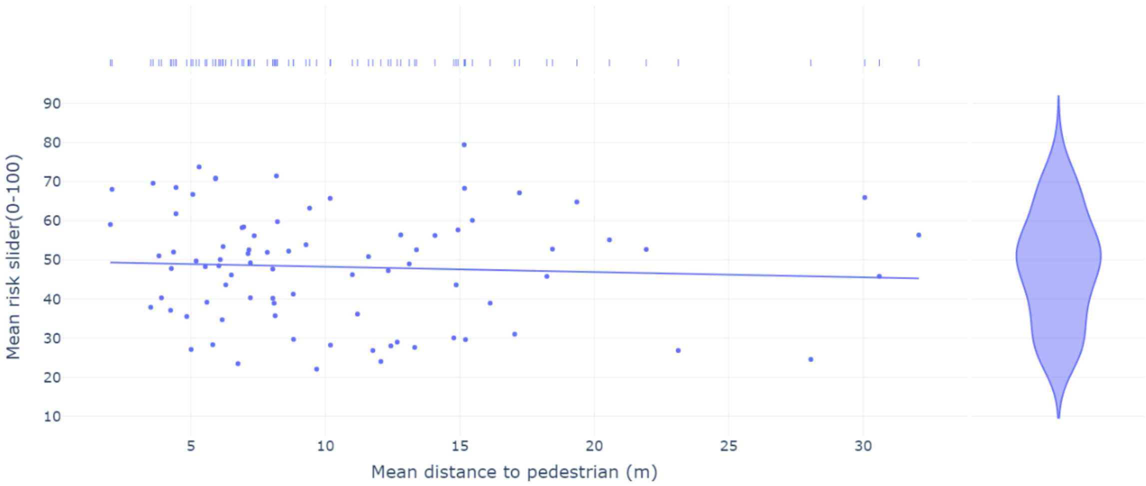

Figure A2. Scatterplot of the post-trial risk slider vs mean distance to the pedestrian ($r = -0.07$). Each dot represents one video ($N = 85$).

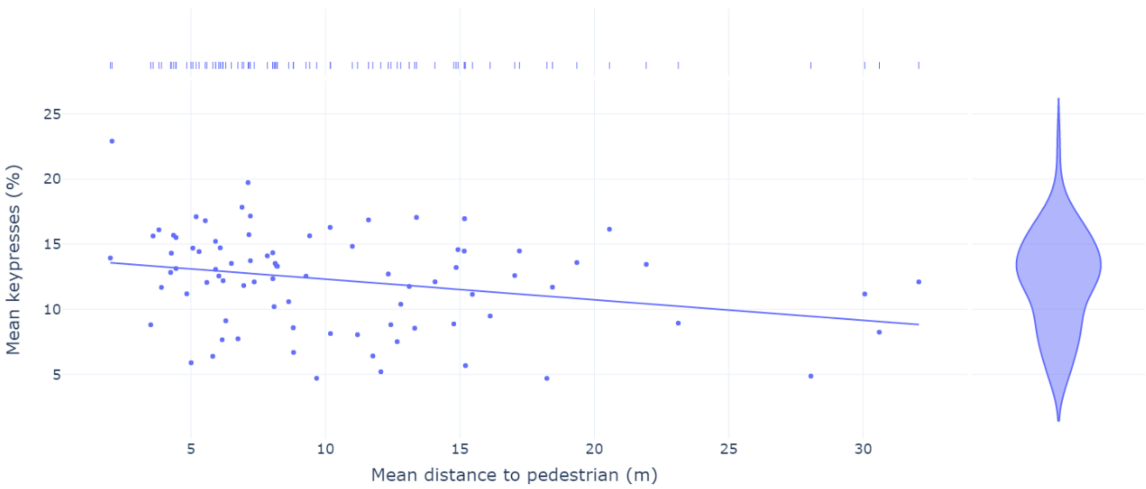

Figure A3. Scatterplot of the mean keypresses vs mean distance to the pedestrian ($r = -0.26$). Each dot represents one video ($N = 85$).

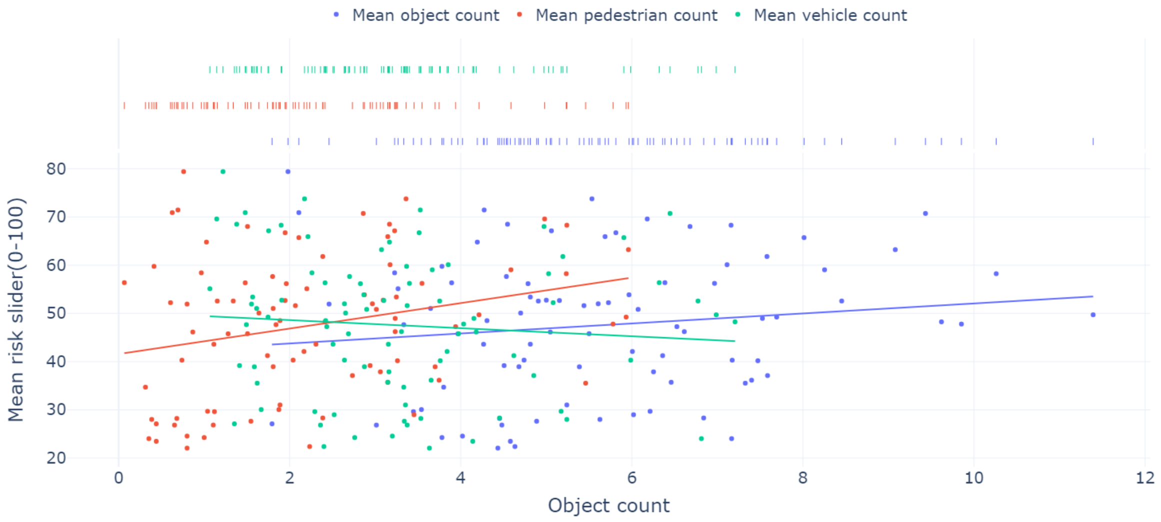

Figure A4. Scatterplot of the mean risk slider values vs mean object count. Each dot represents one video (N = 85). The r values for object count, pedestrian count and vehicle count were respectively: r = 0.14, r = 0.26, r = -0.08.

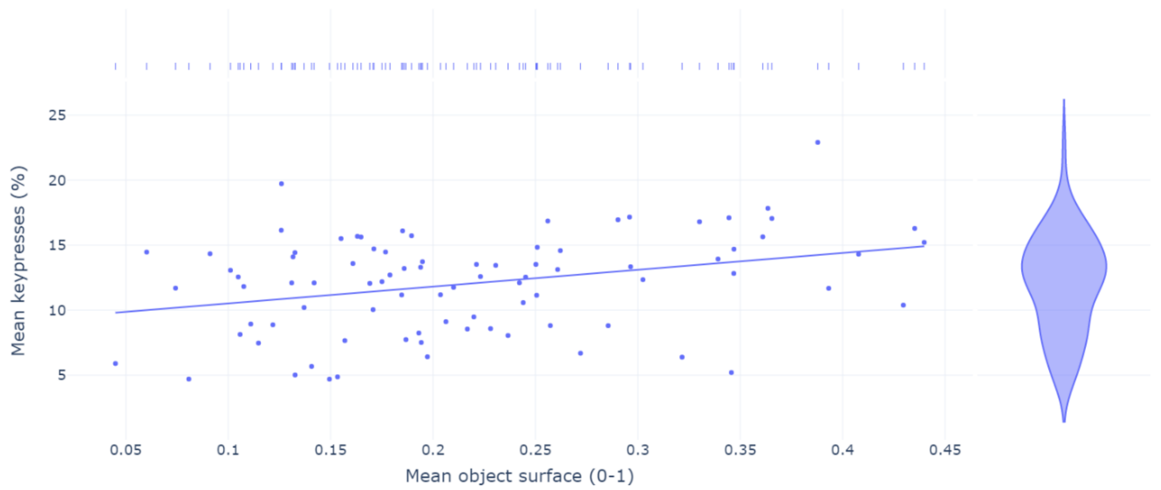

Figure A5. Scatterplot of the mean keypresses vs mean object surface (r = 0.30). Each dot represents one video (N = 85).

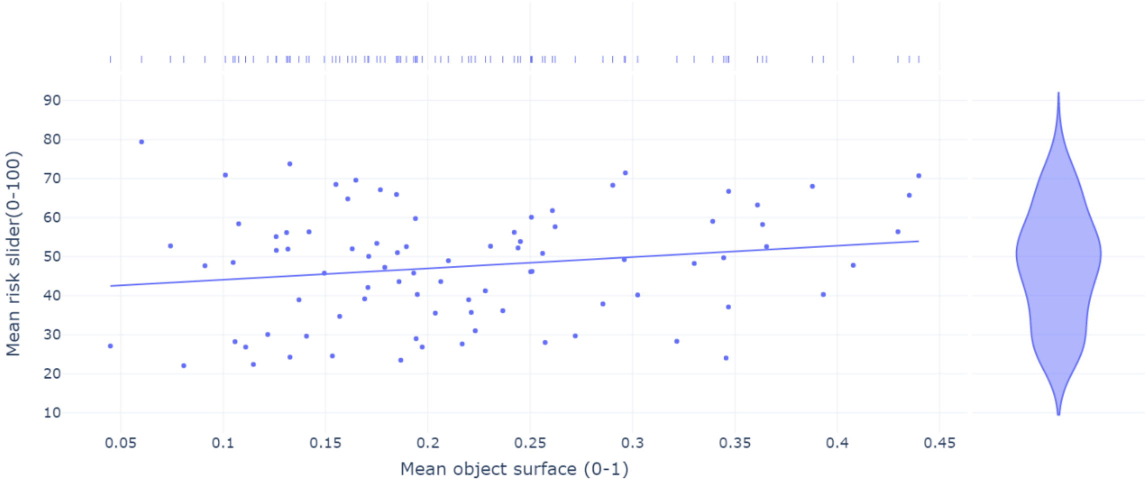

Figure A6. Scatterplot of the mean keypresses vs mean object surface (r = 0.19). Each dot represents one video (N = 85).

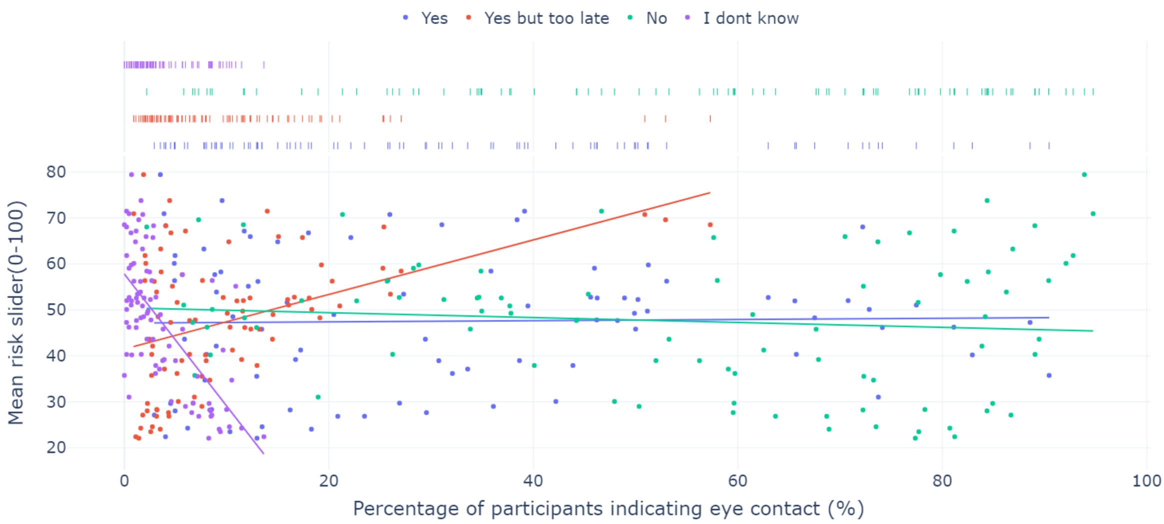

Figure A7. Regression analysis of the mean keypresses versus the percentage of participants indicating each post-trial eye contact answer. Each dot represents one video, with the mean value of the post-trial risk slider (averaged over all participants) on the y-axis, and the frequency of subjective eye contact answers on the x-axis. The following correlations were found: "*Yes*" (r = 0.03), "*Yes, but too late*" (r = 0.42), "*No*" (r = 0.07), and "*I don't know*" (r = -0.65). N = 86 videos were used for this regression.

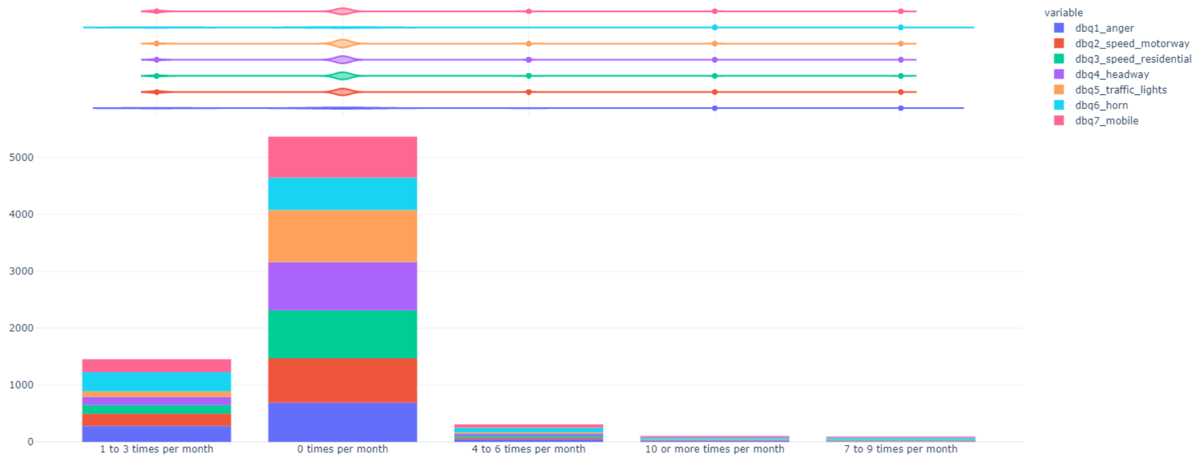

Figure A8. Answers on behavioural questions in the pre-survey questionnaire.

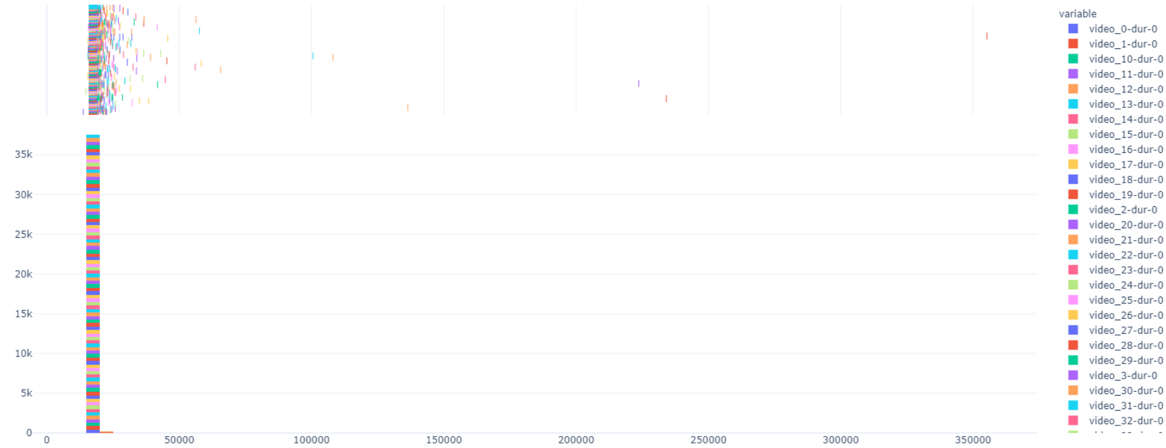

Figure A9. Average duration of videos showed to the participants.

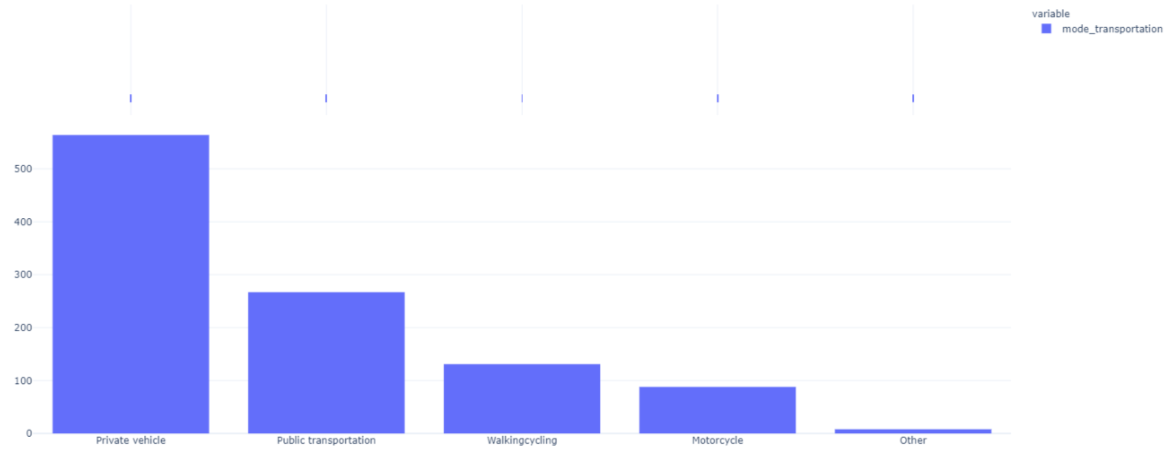

Figure A10. frequency of most common mode of transportation used by the participants.

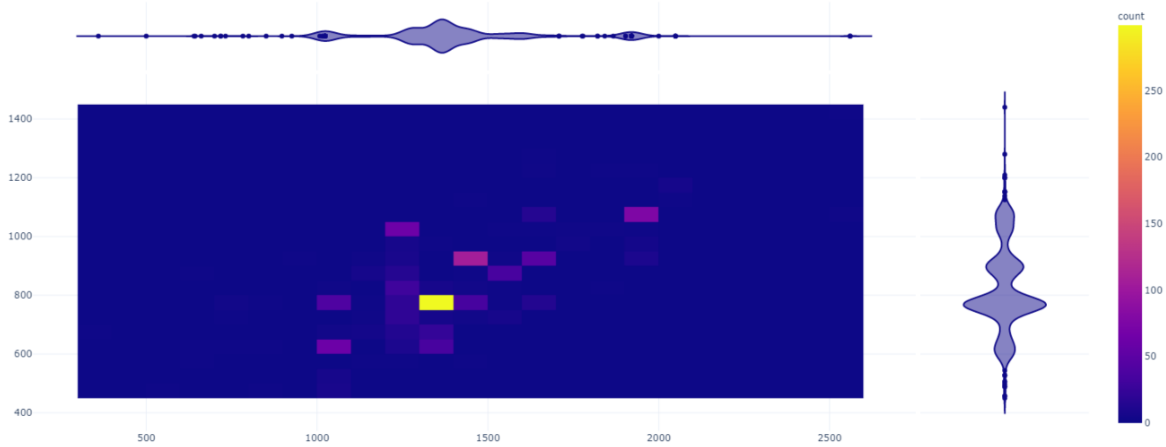

Figure A11. Plot indicating window heights and window widths of the participants.

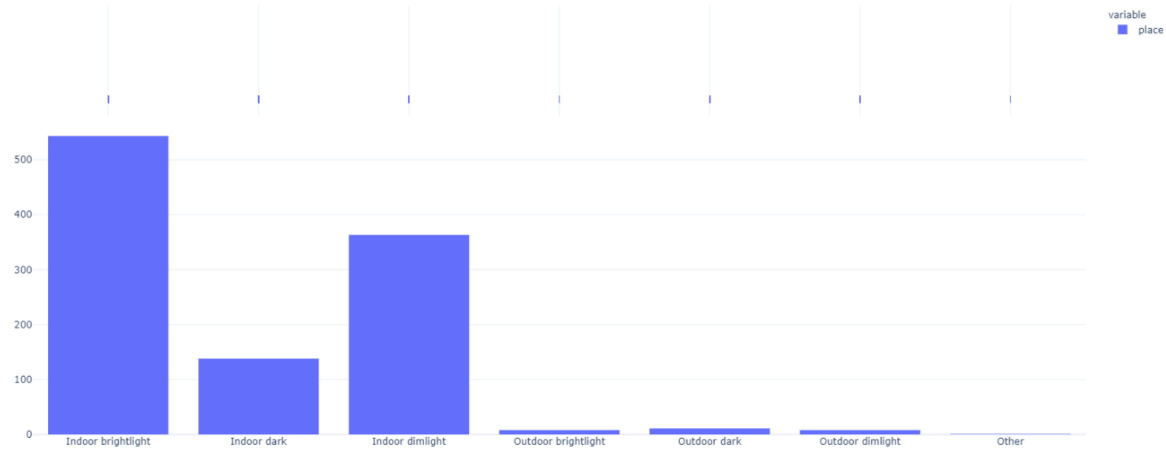

Figure A12. Frequency of location characteristics in which the participant conducted the survey.

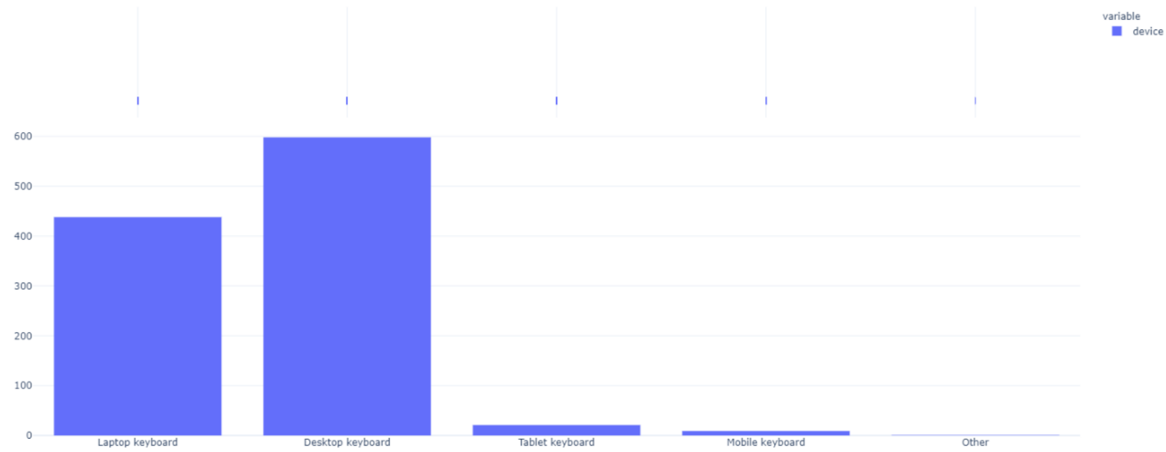

Figure A13. Frequency of each device used to conduct the survey with.

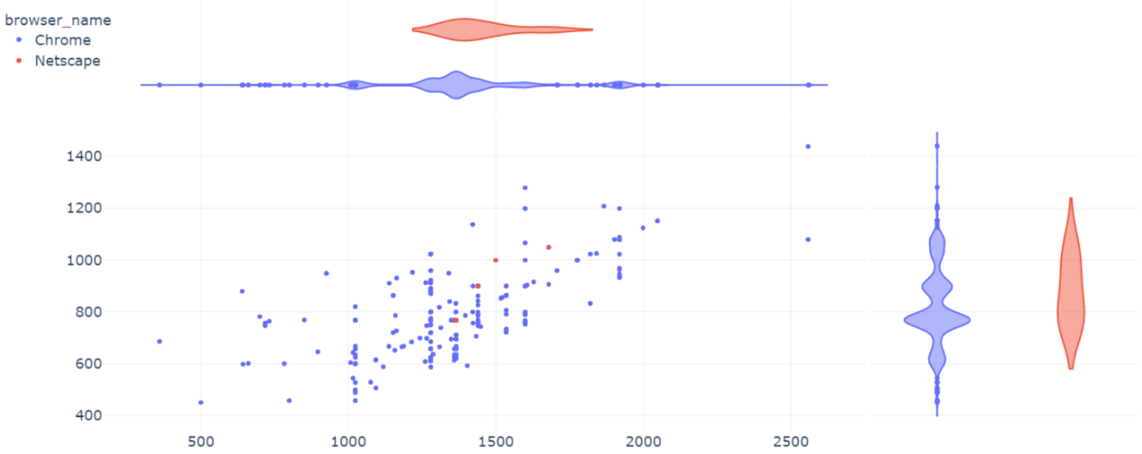

Figure A14. Scatterplot indicating which browser was used for the survey, and screen width and height.

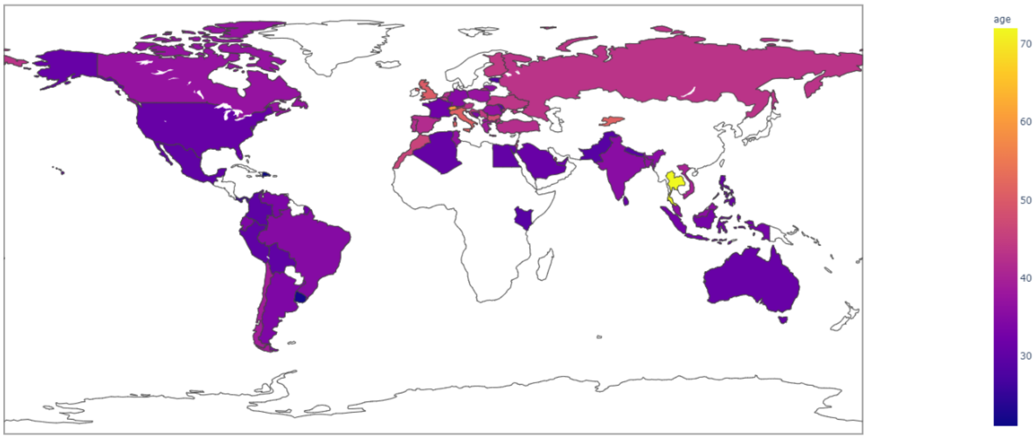

Figure A15. Average age of the participants per country.

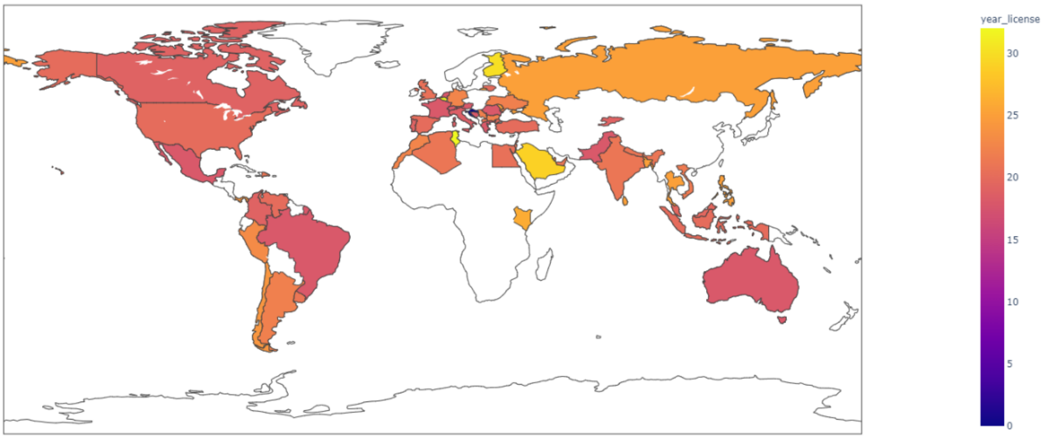

Figure A16. Average year driver license of the participant was achieved per country.

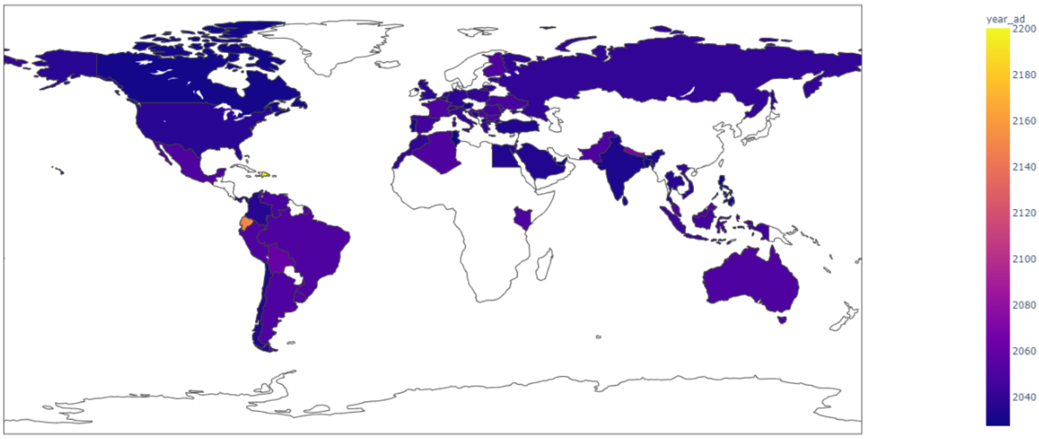

Figure A17. Average year fully autonomous driving is expected per country.

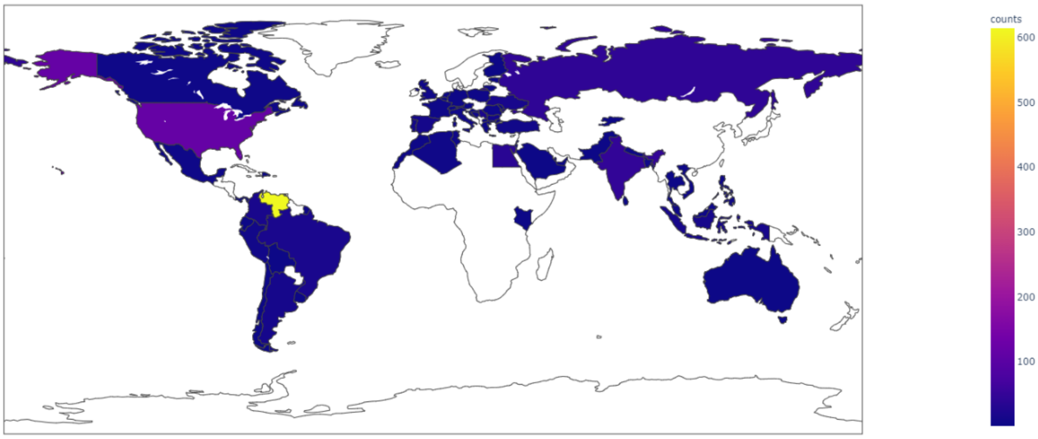

Figure A18. Participant count per country.

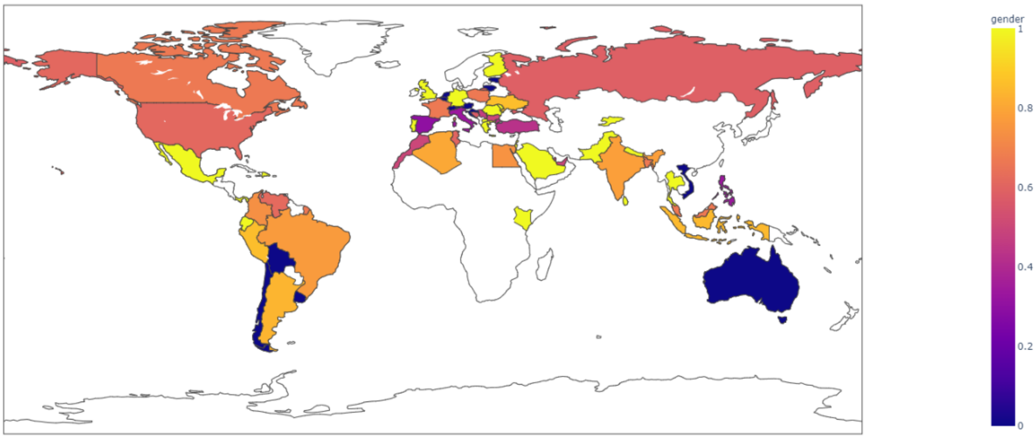

Figure A19. Gender ratio per country. 1 indicates male, whereas 0 indicates female.

Appendix B – Google Earth and Google Cloud Video Intelligence API

Figure B1. Map of Toronto in Google Earth. All blue pins represent the position of the pedestrian where the crossing took place (Google, 2001).

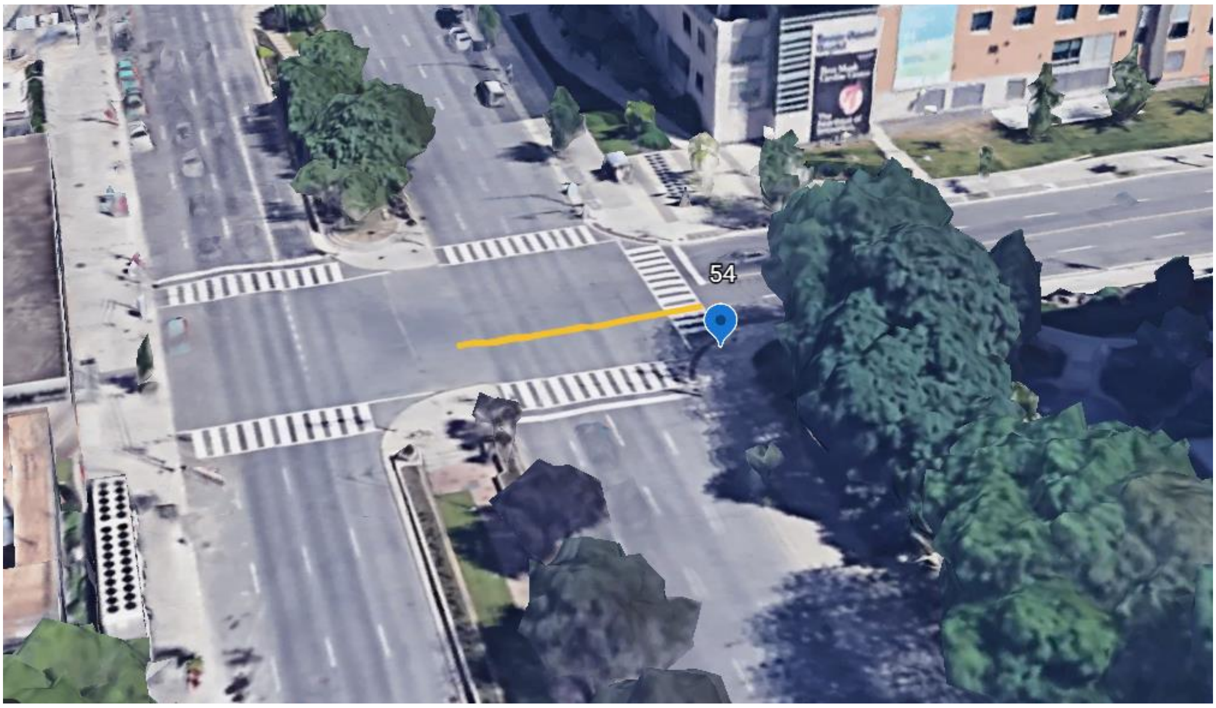

Figure B2. A zoom in on one of the interaction sites in Google Earth. The blue pin represents the location of the pedestrian, whereas the yellow line indicates the route of the vehicle, through annotating the position of the vehicle at the 9th, 10th and 11th second in the video (Google, 2001).

Figure B3. The vehicle dashcam image on the 10th second in the video, of the situation depicted in Figure B2.

```
Processing video for object annotations.

Finished processing.

Entity description: helicopter
Entity id: /m/09ct_
Segment: 0.0s to 4.8s
Confidence: 0.938391923904419
Time offset of the first frame: 0.0s
Bounding box position:
    left  : 0.6771026253700256
    top   : 0.25470441579818726
    right : 0.812613308429718
    bottom: 0.36869969964027405
```

Figure B4. Example of the output when an entity is found in a video, using the object detection algorithm of the Google Cloud Video Intelligence API. (Adil Khan, n.d.)

Appendix C – Appen survey

The following images show the pre-survey questionnaire, which elaborates on the purpose of the survey and includes questions on demographics, behaviour in traffic, and thoughts on eye contact and communication in traffic.

Pedestrian Crossing

Instructions ▴

You are invited to participate in a research study entitled "Risk in situations with pedestrians crossing." The study is being conducted by Bram Kooijman, Dr. Pavlo Bazilinskyy, Dr. Dimitra Dodou, and Dr. ir. Joost de Winter, Department of Cognitive Robotics, Delft University of Technology, The Netherlands. Contact: b.kooijman@student.tudelft.nl (mailto:b.kooijman@student.tudelft.nl).

The purpose of this research is to determine the risk in situations with pedestrians crossing.

You are free to contact the investigator at the above email address to ask questions about the study. You must be at least 18 years old to participate. The survey will take approximately 30 minutes of your time. In case you participated in a previous survey of one of the present researchers, your responses may be combined with the previous survey. The information collected in the survey is anonymous. Participants will not be personally identifiable in any research papers arising from this study.

If you agree to participate and understand that your participation is voluntary, then continue. If you would not like to participate, then please close this page. Before the study starts, the videos will be preloaded. This may take a few minutes depending on your internet connection.

General questions

Have you read and understood the above instructions? (required)

- ○ Yes
- ○ No

What is your gender? (required)

- ○ Male
- ○ Female
- ○ I prefer not to respond

What is your age? (required)

In which type of place are you located now? (required)

- ○ Indoor, dark
- ○ Indoor, dim light
- ○ Indoor, bright light
- ○ Outdoor, dark
- ○ Outdoor, dim light
- ○ Outdoor, bright light
- ○ Other

○ I prefer not to respond

If you answered 'Other' in the previous question, please describe the place where you are located now below.

Which input device are you using now? (required)

○ Laptop keyboard
○ Desktop keyboard
○ Tablet on-screen keyboard
○ Mobile phone on-screen keyboard
○ Other
○ I prefer not to respond

If you answered 'Other' in the previous question, please describe your input device below.

At which age did you obtain your first license for driving a car or motorcycle?

What is your primary mode of transportation? (required)

○ Private vehicle
○ Public transportation
○ Motorcycle
○ Walking/Cycling
○ Other
○ I prefer not to respond

On average, how often did you drive a vehicle in the last 12 months? (required)

○ Every day
○ 4 to 6 days a week
○ 1 to 3 days a week
○ Once a month to once a week
○ Less than once a month
○ Never
○ I prefer not to respond

About how many kilometers (miles) did you drive in the last 12 months? (required)

○ 0 km / mi
○ 1 - 1,000 km (1 - 621 mi)
○ 1,001 - 5,000 km (622 - 3,107 mi)
○ 5,001 - 15,000 km (3,108 - 9,321 mi)
○ 15,001 - 20,000 km (9,322 - 12,427 mi)
○ 20,001 - 25,000 km (12,428 - 15,534 mi)
○ 25,001 - 35,000 km (15,535 - 21,748 mi)

○ 35,001 - 50,000 km (21,749 - 31,069 mi)
○ 50,001 - 100,000 km (31,070 - 62,137 mi)
○ More than 100,000 km (more than 62,137 mi)
○ I prefer not to respond

How many accidents were you involved in when driving a car in the last 3 years? (please include all accidents, regardless of how they were caused, how slight they were, or where they happened) (required)

○ 0
○ 1
○ 2
○ 3
○ 4
○ 5
○ More than 5
○ I prefer not to respond

How often do you do the following?: Becoming angered by a particular type of driver, and indicate your hostility by whatever means you can. (required)

○ 0 times per month
○ 1 to 3 times per month
○ 4 to 6 times per month
○ 7 to 9 times per month
○ 10 or more times per month
○ I prefer not to respond

How often do you do the following?: Disregarding the speed limit on a motorway. (required)

○ 0 times per month
○ 1 to 3 times per month
○ 4 to 6 times per month
○ 7 to 9 times per month
○ 10 or more times per month
○ I prefer not to respond

How often do you do the following?: Disregarding the speed limit on a residential road. (required)

○ 0 times per month
○ 1 to 3 times per month
○ 4 to 6 times per month
○ 7 to 9 times per month
○ 10 or more times per month
○ I prefer not to respond

How often do you do the following?: Driving so close to the car in front that it would be difficult to stop in an emergency. (required)

○ 0 times per month
○ 1 to 3 times per month

- 4 to 6 times per month
- 7 to 9 times per month
- 10 or more times per month
- I prefer not to respond

How often do you do the following?: Racing away from traffic lights with the intention of beating the driver next to you. (required)

- 0 times per month
- 1 to 3 times per month
- 4 to 6 times per month
- 7 to 9 times per month
- 10 or more times per month
- I prefer not to respond

How often do you do the following?: Sounding your horn to indicate your annoyance with another road user. (required)

- 0 times per month
- 1 to 3 times per month
- 4 to 6 times per month
- 7 to 9 times per month
- 10 or more times per month
- I prefer not to respond

How often do you do the following?: Using a mobile phone without a hands free kit. (required)

- 0 times per month
- 1 to 3 times per month
- 4 to 6 times per month
- 7 to 9 times per month
- 10 or more times per month
- I prefer not to respond

How do you feel about the following?: Communication between driver and pedestrian is important for road safety. (required)

- Completely disagree
- Disagree
- Neither disagree nor agree
- Agree
- Completely agree
- I prefer not to respond

As a driver, what does it mean to you when a pedestrian makes eye contact with you? (required)

- I should stop
- The pedestrian should stop
- Both should stop
- Neither should stop
- I prefer not to respond

As a pedestrian, what does it mean to you when a driver makes eye contact with you? (required)

- ○ I should stop
- ○ The driver should stop
- ○ Both should stop
- ○ Neither should stop
- ○ I prefer not to respond

Experiment
You will be asked to leave Appen to participate in the experment. You will need to open the link below. Do not close this tab. In the end of the experiment you will be given a code to input in the next question on this tab. Please take a note of the code. Without the code, you will not be able to receive money for your participation. All videos will be downloaded before the start of the experiment. It may take a few minutes. Please do not close your browser during that time.

Open **this link (https://crossing-crowdsourced.herokuapp.com)** to start the experiment.

Type the code that you received at the end of the experiment. (required)

Miscellaneous questions

In which year do you think that most cars will be able to drive fully automatically in your country of residence? (required)

Please provide any suggestions that could help engineers to build safe and enjoyable automated cars.

Test Validators

Appendix D – Heroku crowdsourcing survey

The following pictures are screenshots of instructions, videos and questions included in the crowdsource survey, as a supplement to what was discussed within the paper.

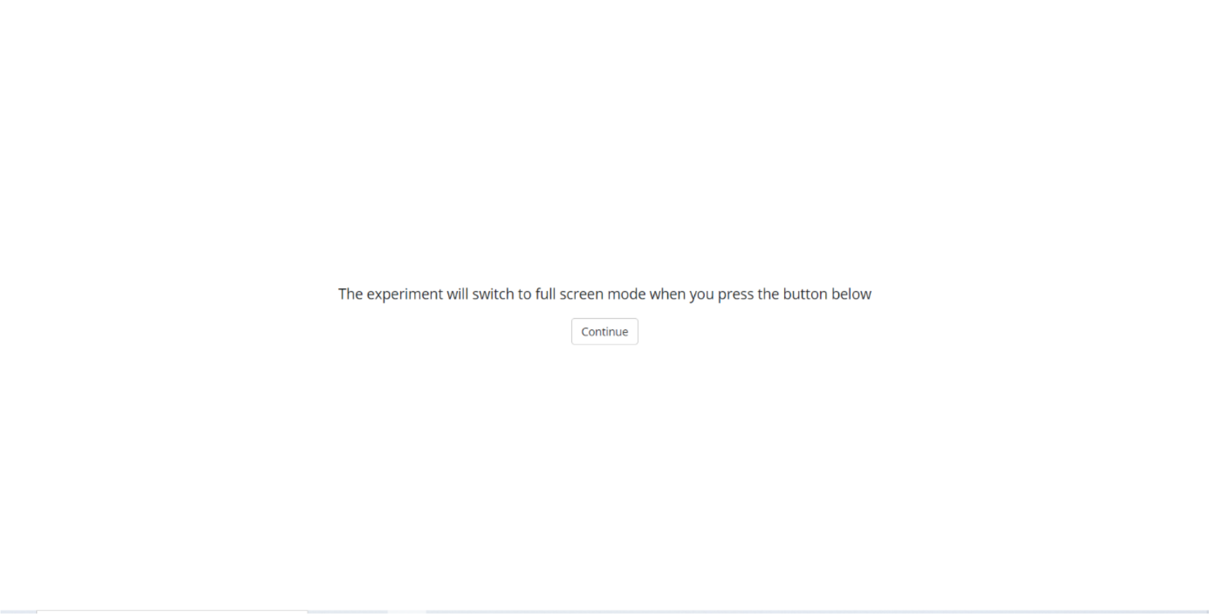

Figure D1. Introductory screen, activating full screen mode for the crowdsourcing survey

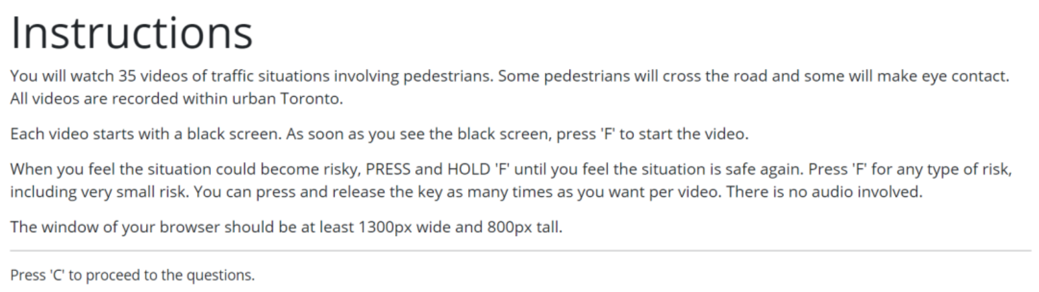

Figure D2. Instructions for the participants regarding the the task they receive, and the videos they will see in this crowdsourcing survey.

Figure D3. Screen that the participants saw before the start of each video

Figure D4. Example of the 1 s blackscreen at the start of the video. Task details are provided below the video

Figure D5. Screenshot of one of the videos that was investigated by the participants, inlucding task details below the video.

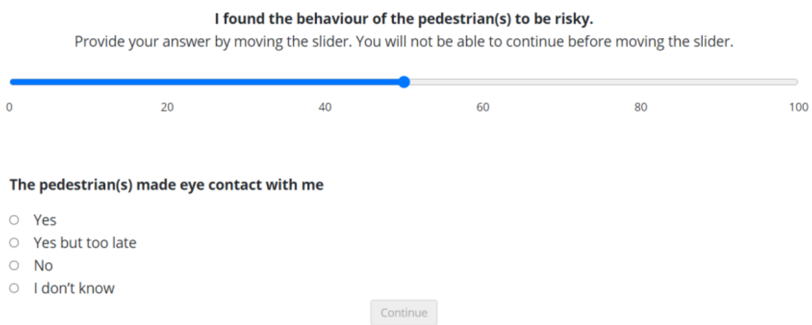

Figure D6. The post-trial questionnaire the participants had to respond to after each video.

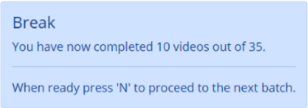

Figure D7. After 10 trials, the participant was presented with the screen shown in this figure.

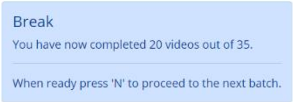

Figure D8. After 20 trials, the participant was presented with the screen shown in this figure.

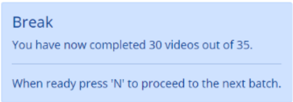

Figure D9. After 30 trials, the participant was presented with the screen shown in this figure.

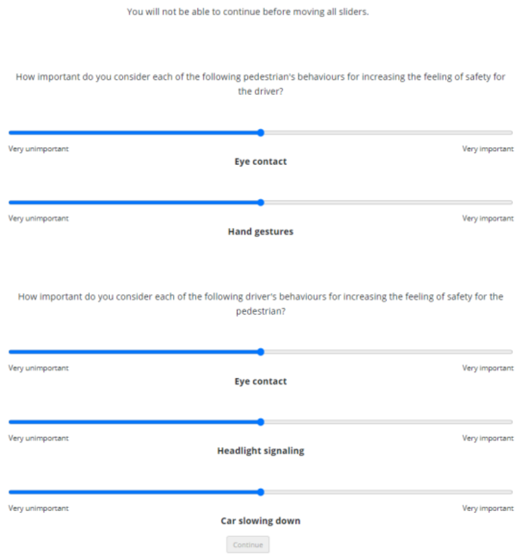

Figure D10. The post-survey questionnaire, which was presented to the participants after watching all 35 videos, regarding driver and pedestrian behaviour.

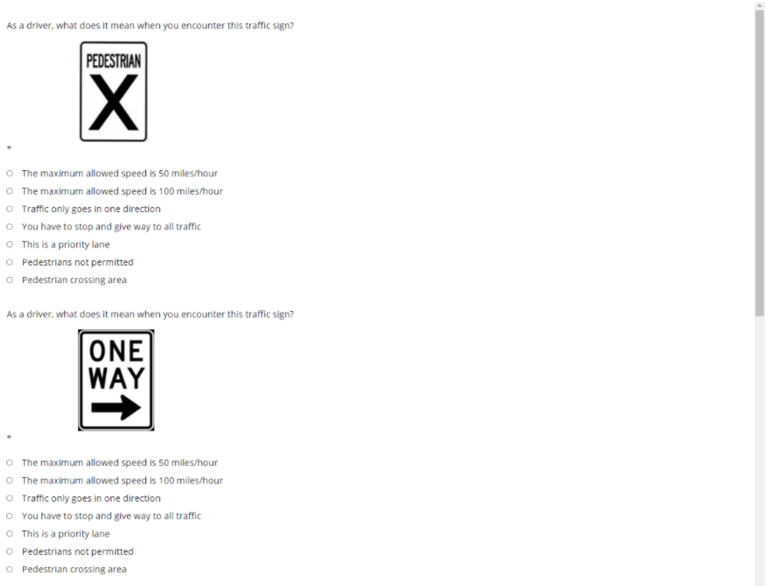

Figure D11. Post-survey questionnaire in which the participant had to answer questions on the meaning of certain traffic signs, as a test of traffic knowledge (pt 1).


As a driver, what does it mean when you encounter this traffic sign?

MAXIMUM
100

*

- The maximum allowed speed is 50 miles/hour
- The maximum allowed speed is 100 miles/hour
- Traffic only goes in one direction
- You have to stop and give way to all traffic
- This is a priority lane
- Pedestrians not permitted
- Pedestrian crossing area

As a driver, what does it mean when you encounter this traffic sign?

STOP

*

- The maximum allowed speed is 50 miles/hour
- The maximum allowed speed is 100 miles/hour
- Traffic only goes in one direction
- You have to stop and give way to all traffic
- This is a priority lane
- Pedestrians not permitted
- Pedestrian crossing area

Continue

Figure D12. Post-survey questionnaire in which the participant had to answer questions on the meaning of certain traffic signs, as a test of traffic knowledge (pt 2).


You have now completed the experiment. Thank you for your participation. Please note down or copy to your clipboard the following code:

X61630935741175BV44601W

Do not forget to enter the code we gave in the Appen page in the field "Type code that you received after the end of the test." to finalise the job and receive your money. You will find this field under the link you clicked to open this tab. You may now close this tab.

Figure D13. After the survey, the participants were presented with this screen, which included a code as proof of participation to acquire the promised money.