Transportation Research Part F: Psychology and Behaviour 93 (2023) 235–247

Contents lists available at ScienceDirect

Transportation Research Part F: Psychology and Behaviour

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

Predicting perceived risk of traffic scenes using computer vision

Joost de Winter [a,*], Jim Hoogmoed [a], Jork Stapel [a], Dimitra Dodou [a], Pavlo Bazilinskyy [a,b]

[a] *Faculty of Mechanical Engineering, Delft University of Technology, the Netherlands*
[b] *Department of Industrial Design, Eindhoven University of Technology, the Netherlands*

A B S T R A C T

Perceived risk, or subjective risk, is an important concept in the field of traffic psychology and automated driving. In this paper, we investigate whether perceived risk in images of traffic scenes can be predicted from computer vision features that may also be used by automated vehicles (AVs). We conducted an international crowdsourcing study with 1378 participants, who rated the perceived risk of 100 randomly selected dashcam images on German roads. The population-level perceived risk was found to be statistically reliable, with a split-half reliability of 0.98. We used linear regression analysis to predict ($r = 0.62$) perceived risk from two features obtained with the YOLOv4 computer vision algorithm: the number of people in the scene and the mean size of the bounding boxes surrounding other road users. When the ego-vehicle's speed was added as a predictor variable, the prediction strength increased to $r = 0.75$. Interestingly, the sign of the speed prediction was negative, indicating that a higher vehicle speed was associated with a lower perceived risk. This finding aligns with the principle of self-explaining roads. Our results suggest that computer-vision features and vehicle speed contribute to an accurate prediction of population subjective risk, outperforming the ratings provided by individual participants (mean $r = 0.41$). These findings may have implications for AV development and the modeling of psychological constructs in traffic psychology.

1. Introduction

There is currently significant attention being paid to the development of automated vehicles (AVs). Next to the many technical challenges involved in creating AVs, there are also various human factors challenges (e.g., Hancock et al., 2020; Merat et al., 2019). Most of the human factors research to date has focused on how drivers respond when they have to take over control from the AV (Lu & De Winter, 2015; Morales-Alvarez et al., 2020; Weaver & DeLucia, 2020; Zhang et al., 2019) and on the design of human–machine interfaces (Bengler et al., 2020; Detjen et al., 2021). A smaller but growing body of research is available on driver state and perception when the AV is driving itself and the driver is a monitor of the AV and the driving environment.

Heikoop et al. (2017) measured drivers' vigilance in a platooning AV in a simulator over several hours. Participants reported a reduction in energetic arousal and concentration compared to before the experiment, but they were still able to detect salient events in the environment. Other researchers have quantified driver drowsiness during automated driving using measures such as eyelid closure and posture (e.g., Hecht et al., 2019; Schömig et al., 2015). Several studies have also used a handheld control unit to continuously poll drivers' discomfort with the AV's driving style in a driving simulator (Charlton et al., 2014; Hartwich et al., 2018; Niermann et al., 2021; Radhakrishnan et al., 2020; Rossner & Bullinger, 2020).

Next to driver-state variables such as drowsiness or comfort, perceived (i.e., subjective) risk may be an important variable in automated driving. In manual driving, perceived risk has been identified as a key construct that influences driver actions (Charlton & Starkey, 2016; Näätänen & Summala, 1974; Wilde, 1982, 2013). Determining perceived risk in an AV might be relevant for several

* Corresponding author.
E-mail address: j.c.f.dewinter@tudelft.nl (J. de Winter).

https://doi.org/10.1016/j.trf.2023.01.014
Received 7 August 2022; Received in revised form 3 January 2023; Accepted 19 January 2023

reasons. First, if the AV can determine perceived risk, it can then determine whether it is safe to hand back control to the human driver or not, something that is of relevance to SAE Levels 3 and 4 automation (Cabrall et al., 2020; Kolekar et al., 2021). Second, AV controllers may monitor perceived risk to optimize the driving experience by, for example, slowing down when perceived risk is determined to be high. By taking perceived risk into account, SAE Levels 3–5 AVs might ensure that they drive in a manner acceptable to drivers, as pointed out by several authors (Guo et al., 2017; Kolekar et al., 2020; Rossner & Bullinger, 2020; Tan et al., 2022; Zhao et al., 2020).

Several studies have explored the relationship between perceived risk and driving conditions. One common method in hazard perception studies is to show observers video clips and ask them to press a key when they detect a hazard (Chapman & Underwood, 1998; McKenna & Crick, 1997; Vlakveld, 2014; for a review, see Moran et al., 2019). Bazilinskyy, Eisma et al. (2020) and Bazilinskyy, Dodou et al. (2020) used this method to measure the perceived risk of participants in different countries. Participants watched dashcam video clips and pressed a key on their keyboard when they perceived a hazard. The results showed that perceived risk varied considerably during the drive, with mostly uneventful scenarios followed by bursts of risk, for example, when approaching pedestrians or other cars. Other studies have found that perceived risk is affected by factors such as visibility (e.g., night vs daytime driving; Evans et al., 2020), weather (e.g., fog vs clear weather; Chen et al., 2019; Saffarian et al., 2012), distance to the car in front (Kondoh et al., 2008; Lewis-Evans et al., 2010; Saffarian et al., 2012; Siebert et al., 2014), lane width (Charlton & Starkey, 2016), and proximity to objects or other road users (He et al., 2022; Kolekar et al., 2020; Kovácsová et al., 2019).

While these studies provide useful information about factors that contribute to perceived risk, they only focus on *measuring* risk, often in pre-programmed scenarios in a driving simulator. It would be valuable to be able to *predict* perceived risk based solely on vehicle state and camera images, i.e., sensors available in AVs. Several studies have attempted to do this. Yurtsever et al. (2019) predicted perceived risk from video clips depicting lane changes. They first applied semantic segmentation masks to the elements in the scene and then fed video frames overlaid with these masks into a Long Short-Term Memory (LSTM) architecture. Ping et al. (2018) used a similar approach, asking human annotators to indicate the risk of individual frames of dashcam videos. They then used an LSTM method on various features, including ego-vehicle state, information on the leading vehicle (e.g., brake lights, position), and the presence of oncoming vehicles and pedestrians. Kolekar et al. (2021) determined drivers' perceived risk based on their verbal utterances and whether they reclaimed control when driving an AV on a test track. The driver's perceived risk was calculated based on whether objects or road boundaries entered a risk field in front of the ego-vehicle, similar to Gibson and Crooks' (1938) 'field of safe travel.' Cabrall et al. (2020) asked participants to report the perceived effort required to reclaim control from the AV after watching short dashcam video clips. They found that the effort ratings were predictable from image characteristics such as road curvature and the presence of other road users. Finally, Bustos et al. (2021) rated perceived stress in dashcam video frames and used various computer vision models to predict the stress ratings from image characteristics. They reported satisfactory prediction accuracy for three computer vision methods: automated scene segmentation, end-to-end image classification, and end-to-end video classification.

However, the above methods have some limitations. For example, Yurtsever et al. (2019) only used overtaking maneuvers and the method of Ping et al. (2018) was partially subjective because the authors manually removed objects that were assumed to be non-risky (such as parked vehicles and pedestrians not along the route of the vehicle), while Cabrall et al. (2020) labeled video frames manually. Additionally, these studies treated risk measurement as a typical labeling/annotation task in computer vision research, where a small number of participants rated video clips (Yurtsever et al., 2019, 10 raters), video frames (Ping et al., 2018, 14 raters; Bustos et al., 2021, 9 raters), or driven road segments (Kolekar et al., 2021, 8 test drivers). These studies gave little consideration to psychometric principles, such as statistical reliability, which is a necessary condition for validity (Nunnally, 1967). Using a small number of subjects (or a lack of repeated measures) may cause predictive models to be less accurate than they should be (Liu & Salvendy, 2009; Rushton et al., 1983).

The goal of this study was to determine whether perceived risk, as rated by a large number of human participants, could be predicted from basic features obtained using computer vision, which would also be available to an AV. To accomplish this, we measured participants' perceived risk for a selection of diverse traffic scenes depicted in photos. While photos do not provide information about speed, they have several advantages over videos. In a video, it may be unclear whether risk perception should be considered excluding or including the driver's response. For example, if a driver in a video approaches a busy traffic situation and brakes, risk could be judged as low (because the driver appropriately slowed down) or as high (because a busy traffic situation is ahead). Using images may provide a more stimulus-driven way of assessing perceived risk.

2. Methods

Images were taken from the KITTI dataset (Geiger et al., 2013), consisting of regular driving of a Volkswagen Station vehicle in the Karlsruhe area, Germany. We retained 210 images from 26 drives in the KITTI dataset, i.e., short video segments between 8 and 80 s long, driven on 26 September 2011 between 13:02 and 15:19. Seventy images of 12 drives were taken from the category City, 70 images of 8 drives from the category Residential, and 70 images of 6 drives from the category Road (including highways and rural roads). The images were taken using uniform sampling in time, with time intervals between images of a drive being 5.0 s, 5.6 s, and 2.6 s for the City, Residential, and Road categories, respectively. All images had a resolution of 1242 $\times$ 375 pixels.

To gather risk ratings for each image, we used a crowdsourcing approach and allowed 1918 contributors from all countries to participate. We did not impose any limitations on the driving experience of our participants but conducted a subgroup analysis to examine whether driving frequency was a moderating factor. The study was approved by the Human Research Ethics Committee of the TU Delft.

Participants enrolled in the study through the crowdsourcing platform Appen. At the top of the Appen page, contact information was provided, and the purpose of the study was described as "*to determine perceived risk of dash camera images. Your participation in this study may contribute to a better understanding of risk and threat perception while driving, and the creation of human threat perception models*." Participants were informed that they could contact the investigators to ask questions about the study and that they had to be at least 18 years old. Information about anonymity and voluntary participation was also provided. Participants first answered demographic questions, such as about their age, gender, and driving experience.

Participants were then directed to the experiment by clicking on a link that opened a webpage created using jsPsych (De Leeuw, 2015). The page presented them with the following instructions: "*The purpose of this research is to determine perceived risk of dash camera images. You will view 100 images. Each image will be on a separate page. For each image, you will need to answer the same question by moving a slider. To advance to the next image, the slider needs to be moved.*" Next, the participants responded to a random subset of 100 of the 210 images, which is the same number of images used in a previous image-based crowdsourcing study that yielded high participant satisfaction ratings (Bazilinskyy, Eisma et al., 2020, Bazilinskyy, Dodou et al., 2020). Below each image, it was mentioned: "*As a driver, how risky would you judge this situation (0 = no risk, 10 = extreme risk)?*" No additional information about the images, such as their locations, was provided.

An example of an image to be rated is shown in Fig. 1. In our study, the perceived risk was multiplied by 10, so that the minimum possible rating was 0% and the maximum possible rating was 100%. After rating 100 images, participants received a code which they entered on the Appen page to redeem their payment. Participants received a payment of USD 0.25 for their participation in the study. We offered this low payment because the crowdsourcing study could be completed from any country in the world, including low-income countries, and we did not want the financial remuneration to be coercive.

Computer vision was used to detect objects in the images. Instead of using the manually labeled bounding boxes from the KITTI dataset, we decided to use a widely available method for automatic object detection, as this may give a more realistic picture of the capabilities of AVs. Specifically, a YOLOv4 model (Bochkovskiy et al., 2020) pretrained on the COCO dataset (Lin et al., 2014; obtained from sbairagy-MW, 2021) was used to extract instances of objects of a certain class per image (see Fig. 2 for two examples). YOLOv4 can identify a total of 80 different classes. For this analysis, only classes that are related to traffic were retained: Person, Bicycle, Car, Motorbike, Bus, Train, Truck, Traffic light, and Stop sign.

3. Results

3.1. Data filtering

The 1918 respondents took part between November 23 and December 17, 2020. A satisfaction survey offered by Appen yielded the following ratings on a scale of 1 to 5: Instructions clear: 4.2, Ease of job: 3.8, Pay: 3.7 (115 participants completed this optional survey). Participants who indicated they did not read the instructions, completed the study in less than 5 min, provided risk ratings with a standard deviation across images of less than 5% (suggesting not to have taken the task seriously), or yielded ratings for fewer than 90 of the 100 images (due to database storage errors) were removed from the analysis. Furthermore, if a participant completed the study more than once from the same IP address, only the data from the first attempt of that participant was kept. In total, 540 of the 1918 responses were removed, mostly because of duplicate IP addresses, leaving 1378 participants. Their mean age was 36.7 years (SD = 11.4). Of the participants, 895 were male, 477 were female, and 6 preferred not to respond. The mean number of risk ratings available per participant (n = 1378) was 99.5 (SD = 1.52). The mean number of risk ratings available per image (n = 210) was 652.6 (SD = 16.8).

Fig. 1. Measurement of perceived risk using crowdsourcing. Each participant rated a random 100 of the 210 images.

Fig. 2. Results of YOLOv4 for 2 of the 210 images. In the upper image, the algorithm identified six persons and a backpack. In the lower image, the algorithm identified 13 cars and 2 trucks.

3.2. Population perceived risk per image

First, we computed a population perceived risk (PPR) score for each of the 210 images by taking the average rating across the participants. Next, we examined whether the PPR was statistically reliable. Fig. 3 provides a visualization of split-half reliability, that is, the PPR values computed for participants split into two groups. A strong correlation ($r = 0.98$) can be seen, indicating that the PPR was reliable at the level of images.

Further exploration at the level of participants ($n = 1378$) showed no strong associations ($|r| < 0.06$) between participants' mean

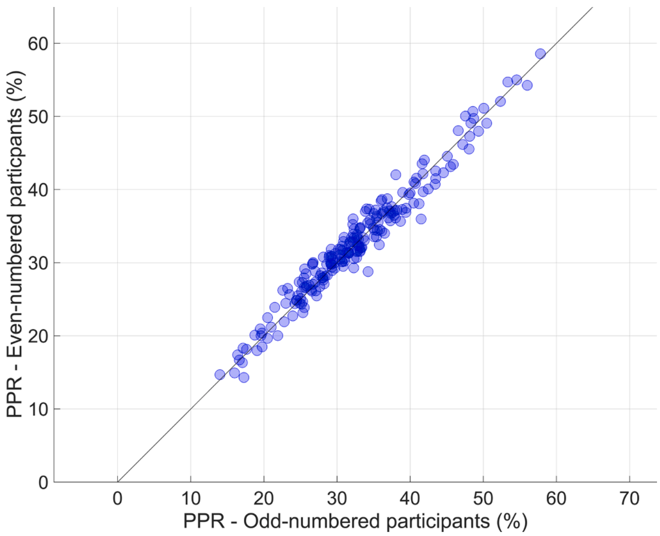

Fig. 3. Mean perceived risk per image as reported by even-numbered participants ($n = 689$) versus mean perceived risk as reported by odd-numbered participants ($n = 689$). Each marker represents an image and is based on an average of 326 responses.

risk across the rated images and gender (1 = female, 2 = male), age, driving mileage in the past 12 months (rated on a scale from 1 = 0 km to 10 = more than 100,000 km), and driving frequency in the past 12 months (rated on a scale from 1 = never to 6 = every day). Fig. 4 shows the PPR calculated from participants who reported driving only rarely in the past 12 months versus participants who reported driving more frequently. The correlation at the level of images is strong ($r = 0.96$), signifying that the frequency of driving had no major influence on the risk ratings.

There were, however, national differences. An examination of perceived risk averaged across all images showed differences in mean risk for participants from Venezuela (32.6%, $n = 591$), the USA (37.8%, $n = 95$), Russia (27.0%, $n = 68$), Ukraine (24.2%, $n = 64$), and India (34.5%, $n = 57$). Fig. 5 shows PPR values divided into participants from the USA ($n = 95$) versus participants from Ukraine and Russia combined ($n = 132$). It can be seen that participants from Ukraine and Russia reported lower risk than participants from the USA, except for the higher-risk images. However, the correlation coefficient is again strong ($r = 0.82$), which is high considering that the PPR values are based on smaller sample sizes than in Figs. 3 and 4. In comparison, according to a bootstrapping analysis with two groups of sizes 95 and 132, respectively, the mean correlation coefficient of the PPR ratings of the two images between the two groups was found to be 0.86 ($SD = 0.019$), barely higher than the $r = 0.82$ association depicted in Fig. 5.

In other words, while there may be national differences in perceived risk (see also Bazilinskyy, Eisma et al., 2020, Bazilinskyy, Dodou et al., 2020; Lee et al., 2020; Nordfjærn et al., 2011; Ventsislavova et al., 2019), the relative differences in PPR values between images seem largely invariant to nationality, as well as age, gender, and driving experience.

3.3. Applying computer vision to the images

Table 1 provides an overview of the output of the YOLOv4 algorithm. The most frequently occurring class per image was Car (3.47). Two features were extracted from the YOLOv4 bounding box results:

(1) *The number of persons (Persons)*. As shown in Table 1, the number of persons in the image is predictive of PPR. Only 15% of the images contained persons, but when they were present, PPR was higher compared to when they were absent. Bicycles were not counted separately since a person is typically on the bicycle.

Psychologically, pedestrians imply an unpredictable environment. Fig. 2 (top panel) shows an image with a large number of persons (pedestrians). Although the ego-vehicle is not on an evident collision course with the pedestrians, and the pedestrians are still far away, it can be imagined that careful driving is advised here since one of the pedestrians (or other road users in the scene) may have overlooked the car and step on the road. More generally, the presence of persons suggests driving in a city environment, where diverse types of events may happen, some of which are hazardous.

(2) The mean square root of the area of the instances of the classes (Area):

$$Area = \frac{1}{n}\sum_{i=1}^{i=n}\sqrt{w_i h_i}$$

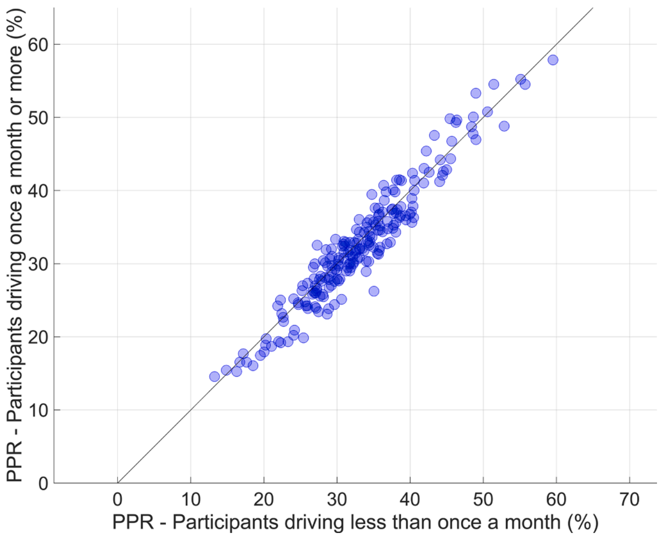

Fig. 4. Mean perceived risk per image for participants who reported driving less than once a month ($n = 281$) versus participants who reported driving once a month or more frequently ($n = 1079$). Each marker represents an image and is based on an average of 133 and 511 responses for the former and latter groups of participants.

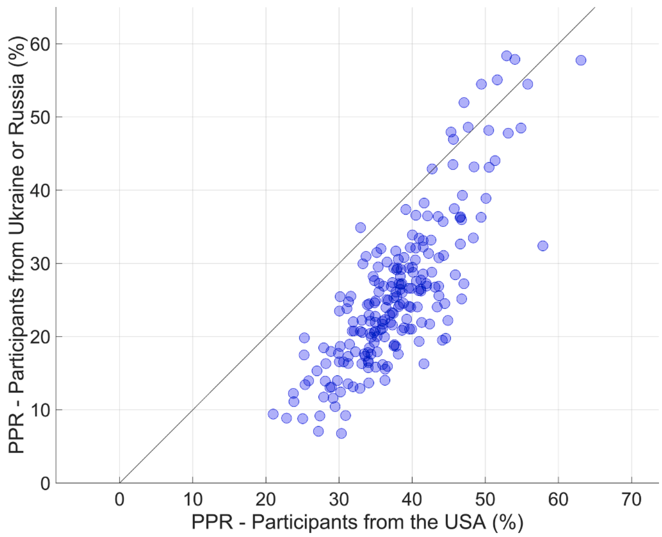

Fig. 5. Mean perceived risk per image for participants from Ukraine and Russia combined ($n = 132$) versus participants from the USA ($n = 95$). Each marker represents an image and is based on an average of 63 and 45 responses for the former and latter groups of participants.

Table 1
Mean and standard deviation (SD) of the number of class instances, proportion of images with the class instance, and Pearson correlation coefficient between the number of class instances and population perceived risk (PPR) ($n = 210$).

	Mean	*SD*	Proportion	*r*
Person	0.27	0.93	0.15	0.33
Bicycle	0.11	0.47	0.08	0.30
Car	3.47	3.24	0.76	0.18
Motorbike	0.01	0.12	0.01	0.06
Bus	0.05	0.24	0.05	0.08
Train	0.05	0.22	0.05	−0.12
Truck	0.35	0.59	0.30	0.31
Traffic light	0.28	0.74	0.16	0.21
Stop sign	0.04	0.19	0.04	0.06

where i is the instance number, n is the number of instances for Persons, Bicycles, Cars, Motorbikes, Buses, Trains, or Trucks, w is the width of the bounding box in pixels, and h is the height of the bounding box in pixels. If there were no instances at all ($n = 0$), then Area was set to 0 px.

Psychologically, the Area measure represents proximity-related risk. Fig. 2 (bottom panel) shows an image with a large number of cars. The sheer number of cars may not be perceived as risky because many of the cars are parked next to the road or in a separate parking lane (see also the relatively low correlation between PPR and the number of cars in Table 1). However, some of the cars are close, as signified by a large bounding box, and may, therefore, contribute to a feeling of risk.

Fig. 6 shows eight images sorted by perceived risk in eight octiles, along with the scores of the two features (Persons and Area), and vehicle speed obtained from the KITTI dataset. The lowest PPR was found for a completely straight and empty road. The highest PPR was found for an image with a car that was approaching from the front, which was reflected in a relatively large Area score.

3.4. Predicting population perceived risk

The correlation matrix (Table 2) shows that the two computer vision features had substantial correlations with PPR. A third measure, vehicle speed, showed a strong negative correlation with PPR. Additionally, the Area measure tended to be smaller when vehicle speed was higher. One possible explanation is that drivers may adopt larger margins with other vehicles on higher-speed roads. Table 2 also shows correlations with the median time it took participants to enter their response, revealing that images perceived as less risky were rated more quickly.

The results of a linear regression analysis using the number of persons and Area as predictors of perceived risk are shown in Table 3. The predictive correlation of the weighted mean of the number of persons and Area equaled $r = 0.62$. Vehicle speed was not available to participants in the crowdsourcing study, but it is a commonly measured variable in manual and automated vehicles. Table 4 shows the results of the regression analysis with the addition of vehicle speed. The predictive correlation of the regression model is $r = 0.75$,

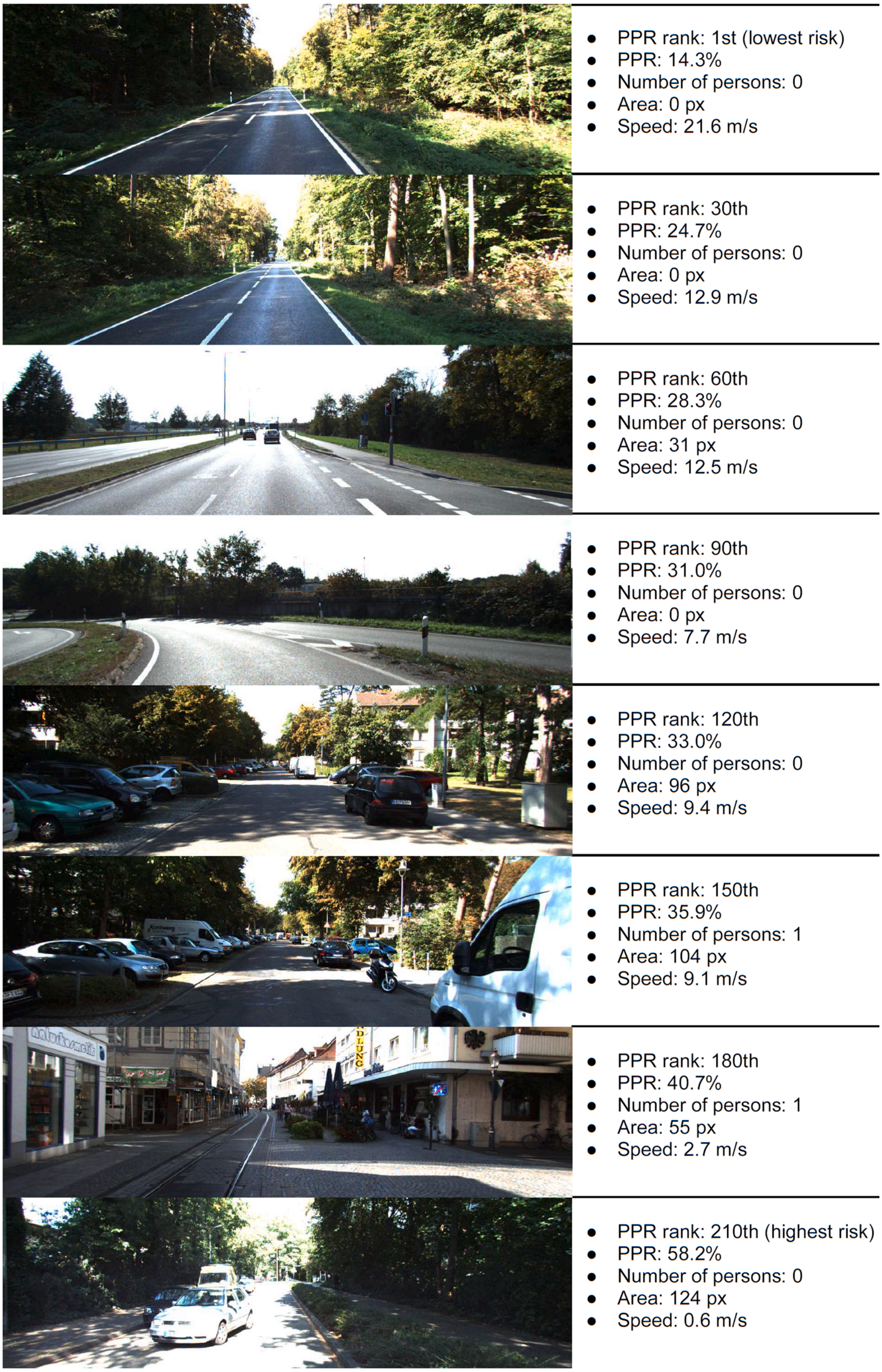

Fig. 6. Eight of the 210 images, sorted from the lowest to the highest perceived risk.

Table 2
Correlation matrix of the two computer-vision features, vehicle speed, population perceived risk (PPR), road type, and population median response time ($n = 210$).

Variable	Mean	*SD*	1	2	3	4	5
1. Number of persons (#)	0.27	0.93					
2. Area (px)	62.77	48.81	0.06				
3. Speed (m/s)	9.05	5.37	−0.10	−0.41			
4. Population perceived risk (%)	32.64	8.09	0.33	0.54	−0.63		
5. Road type (1 = city, 2 = residential, 3 = road)	2.00	0.82	−0.28	−0.21	0.47	−0.32	
6. Population median response time (ms)	4115	374	0.32	0.31	−0.59	0.73	−0.37

Table 3
Regression analysis results for predicting population perceived risk (PPR) from computer-vision variables ($n = 210$).

	Unstandardized *B*	Standardized β	*t*	*p*
Intercept	26.54			
Number of persons	2.61	0.30	5.47	< 0.001
Area (px)	0.09	0.52	9.44	< 0.001

as illustrated in Fig. 7. An example of an image where risk was misestimated is shown in Fig. 8. In this case, participants gave higher risk ratings than the model predicted.

Finally, it was examined how well individual participants could predict the PPR values. The distribution of the predictive correlations for the 1378 participants is shown in Fig. 9. The mean correlation coefficient was 0.41, with 1st, 5th, 25th, 50th, 75th, 95th, and 99th percentiles of −0.24, −0.02, 0.29, 0.45, 0.57, 0.69, and 0.75. In other words, on average, a participant's risk ratings for the 100 rated images correlated with the PPR values for those images at a coefficient of 0.41. These results suggest that computer vision may have the potential to outperform human raters. In particular, the regression model shown in Table 3 had a predictive r value at the 84.9th percentile, while the regression model in Table 4 performed at the 98.6th percentile.

4. Discussion

Automated vehicles (AVs) are designed to navigate safely through traffic, but they might not be attuned to how the human driver inside the vehicle feels. It would be valuable to be able to predict the driver's perceived risk solely based on information obtained from AV sensors. This would allow for optimizing the ride in the AV without relying on complex apparatuses like driver-state monitoring systems.

This study showed that computer vision outcomes can accurately predict population-level perceived risk in still images of traffic scenes. Based on only two features (the number of persons and the mean area of bounding boxes), the predictive accuracy was determined to be $r = 0.59$, and with the addition of speed, the predictive accuracy increased to $r = 0.74$. These findings suggest that basic features computed from camera images combined with standard vehicle-state information (such as speed) can allow for a fairly strong prediction of perceived risk in the population.

The current findings resonate with research that has shown that computer algorithms are able to predict various human behaviors or traits. For example, using a method similar to ours (data reduction followed by regression analysis), Kosinski et al. (2013) found that personal characteristics such as age, intelligence, and scores on a personality test were predictable with considerable accuracy from those individuals' Facebook likes. Also using regression analysis, Nagle and Lavie (2020) found that participants' ratings of the visual complexity of images were predictable ($r = 0.70$) from a number of image statistics such as entropy, edge density, etc. The authors found improved predictive accuracy ($r = 0.83$) when using a deep convolutional network instead that was trained to learn perceived ratings of visual complexity. The automated assessment of images is of particular interest in diagnostic imaging in radiology as well (e.g., Hardy & Harvey, 2020).

The findings of our study may represent an advance in traffic psychology. While perceived risk has previously been measured using

Table 4
Regression analysis results for predicting population perceived risk (PPR) from computer-vision variables, including speed ($n = 210$).

	Unstandardized *B*	Standardized β	*t*	*p*
Intercept	34.90			
Number of persons	2.32	0.27	5.73	< 0.001
Area (px)	0.05	0.33	6.52	< 0.001
Speed (m/s)	−0.70	−0.46	−9.11	< 0.001

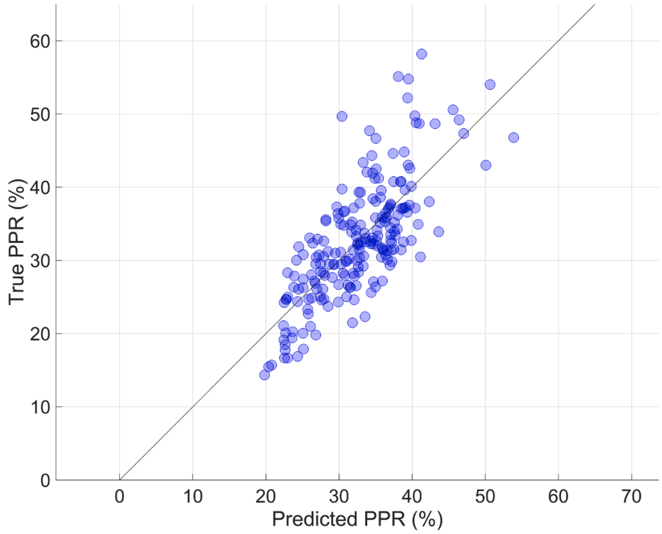

Fig. 7. True perceived risk (i.e., mean perceived risk per image) versus predicted perceived risk (using the regression model depicted in Table 4). Each marker represents an image.

Fig. 8. Image with the largest underestimation of risk (true PPR: 49.7%, predicted PPR: 30.4%).

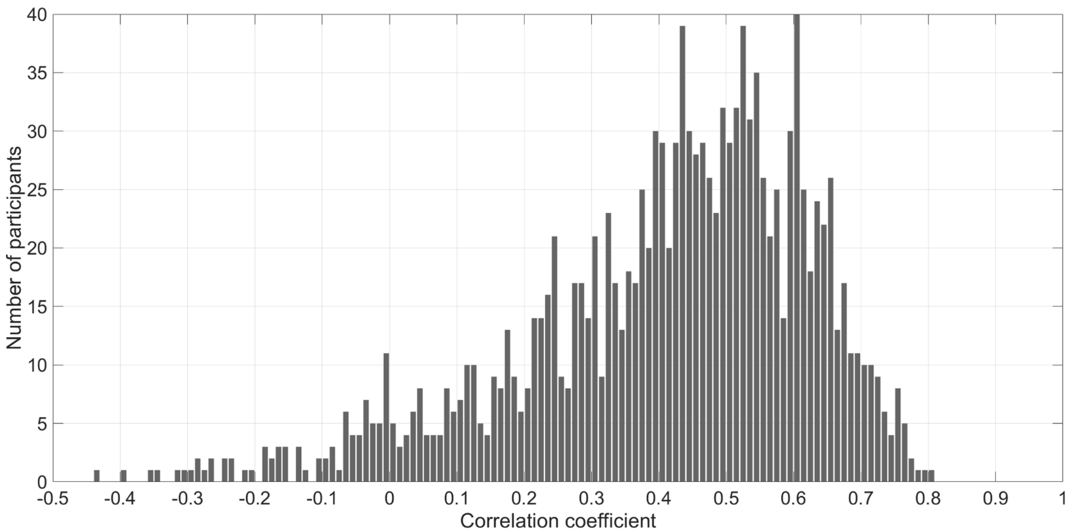

Fig. 9. Distribution of observed correlation coefficients between participants' (n = 1378) perceived risk of the 100 images they rated and the population perceived risk (PPR) of the same images.

button/key presses (e.g., Lee et al., 2020; Ventsislavova et al., 2019), post-trial/drive ratings (Charlton et al., 2014; Pascale et al., 2021), physiological indicators such as skin conductance (e.g., He et al., 2022; Kinnear et al., 2013; Melman et al., 2018; Stapel et al., 2022; Taylor, 1964), fMRI activity (Megías et al., 2018), or questionnaires in which people have to imagine being a driver (e.g., Brell et al., 2019; Hulse et al., 2018), the present findings suggest that perceived risk can be predicted from the driving scene automatically, without relying on human measurements that may be statistically unreliable. The predictive features are considered psychologically plausible. That is, the number of persons as a risk indicator aligns with the fact that driver-pedestrian interaction is a long-known issue in traffic safety, which is partly due to the fact that the behavior of pedestrians can be hard to predict (for human drivers as well as for computer algorithms, Rasouli et al., 2018; Rudenko et al., 2020). The area of the bounding boxes, or proximity, aligns with the idea that driving close to other road users affects perceived risk (Kondoh et al., 2008), similar to Gibson and Crooks (1938), who explained this in terms of one's field of safe travel being impeded.

Although our study used only two computer-vision features and a reasonable number of images (210), there is still a risk of overfitting. For example, persons were included as a strong predictor of perceived risk in the KITTI dataset, which featured shared spaces (see Fig. 2, upper panel). The number of people in the image may be a less powerful predictor in city environments with sidewalks on which many pedestrians may be walking. Future research will have to determine whether the features are similarly predictive in other types of driving environments.

One limitation of this study is that it used still images, for reasons outlined in the introduction. The use of images is supported by Charlton et al. (2014), who found a strong correlation between risk ratings of still images and videos for 12 stimuli (r = 0.93). Nonetheless, previous research suggests that, in some cases, relative speed is a clear predictor of perceived risk. For example, Bazilinskyy, Eisma et al. (2020) and Bazilinskyy, Dodou et al. (2020) found that an unexpected stationary vehicle on a highway was associated with a burst of momentary perceived risk, which would have been impossible to perceive from a still image. Another limitation is that the computer vision algorithm combined with dashcam images does not capture all facets of risk. An example is provided in Fig. 8, where humans rated the risk level as high, while the predicted risk was low because other road users were still quite far away. A possible explanation is the railroad that crossed the main road. Presumably, the driver in the video had already assessed the situation as safe to cross, but the small field-of-view dashcam screenshots may not have conveyed that impression to participants. The inclusion of additional features, such as lead vehicle brake lights as used by Ping et al. (2018), may help to resolve the issue of unexplained risk.

Another limitation of this research was that it was concerned with predicting perceived risk at the population level. It has been argued that perceived risk is person-specific and should therefore be predicted at the individual level (Ping et al., 2018). A problem with predicting risk at the individual level is that human ratings of risk (and other types of self-ratings as well as physiological measurements, etc.) are known to be statistically unreliable (as seen in this study, where the participant-population correlation of perceived risk was low, at 0.41). This makes it practically difficult to determine person-specific reference values. The dilemma of whether risk should be seen as a collective or a person-specific variable relates to the question of whether advanced driver assistance systems (ADAS) should be personalized or whether a one-size-fits-all approach should be taken (e.g., Lefèvre et al., 2015; Van Paassen et al., 2015).

There has also been debate about whether risk measurements in traffic psychology should focus on perceived (subjective) risk, subjective estimates of objective risk, or objective risk itself (such as the probability of collision or loss of control) (Fuller et al., 2008; Kinnear et al., 2008). Additionally, it may be questioned whether perceived risk is distinct enough from other often-used and closely related constructs, such as effort, comfort, and trust (He et al., 2022; Siebert et al., 2014), an issue related to the problem of construct proliferation (Heikoop et al., 2016). It is our contention that we would have obtained similar results if we had used a different rating scale, such as perceived effort, instead of perceived risk (Cabrall et al., 2020).

The use of images may have contributed to the negative correlation between ego-vehicle speed and perceived risk because participants may perceive the risk to be higher when the ego-vehicle was driving slowly. For example, the highest-risk image (Fig. 6, bottom) would likely not be perceived as high-risk if the participant could see from a video that the ego-vehicle was almost stationary. However, the absence of speed information is not the only explanation for the negative correlation between perceived risk and ego-vehicle speed. Previous research using dashcam video clips displayed in a driving simulator also found a strong negative correlation between preferred driving speed and risk ratings (r = -0.625; Charlton & Starkey, 2016), and recent online research using dashcam video clips of pedestrian-crossing situations also found a strong correlation between perceived risk and ego-vehicle speed (r = -0.45, r = -0.52; Kooijman, 2021). One possible explanation is that lower-risk environments, such as straight roads with clear priority rules, permit faster driving (Kooijman, 2021). Another example is provided by the 90th-risky image in Fig. 3: Despite the absence of detected road users, perceived risk is moderately elevated compared to a straight and empty road (lowest risk image). The curve ahead may have caused the ego-driver to slow down, leading to the moderately elevated perceived risk. In other words, a slower speed can be indicative of risk because the ego-driver chose to drive more slowly. It is possible that including more features, such as road curvature (Cabrall et al., 2020), the identification of relevant road users versus irrelevant ones, such as parked cars (Ohn-Bar & Trivedi, 2017; Rashed et al., 2019), or higher-resolution images allowing the computer vision algorithm to detect objects farther away, would reduce the importance of ego-vehicle speed in the regression model.

The above statements relate to the notion of self-explaining roads, which involves the idea that perceptual features of road design, such as lane width, delineation, and amenities for pedestrians, cause drivers to drive faster or slower (Charlton et al., 2010; Van der Horst & Kaptein, 1998). The present findings highlight an interesting paradox, where, normally, for a given situation, a higher speed is associated with higher risk (e.g., Elvik et al., 2019; Fuller et al., 2008; Gårder, 2004; Kinnear et al., 2008; Lewis-Evans & Rothengatter, 2009), but across situations, the opposite is true. It is this paradox that is exploited in the design of shared spaces, following the principles of Monderman (Engwicht, 2012).

The present study demonstrated the potential of computer vision by examining the number of persons, the proximity of objects, and vehicle speed as key predictors. Other research has found that various factors, such as road maintenance, infrastructural issues, brake lights, and overtaking or lateral maneuvers of other vehicles, contribute to perceived risk (Bazilinskyy, Eisma et al., 2020, Bazilinskyy, Dodou et al., 2020; Yurtsever et al., 2019). It is possible that other types of computer vision methods or the use of neural networks on image pixels directly (e.g., Nagle & Lavie, 2020) could improve the prediction of perceived risk. While our study focused on the perceived risk of a traffic situation from the driver's perspective, there are other opportunities to use a more location-specific approach by examining the relationship between (panoramic) images of road environments (e.g., obtained using Google Street View) and objective risk (i.e., crash likelihood; Cai et al., 2022), or to examine how road and city design contribute to perceived risk (Kwon & Cho, 2020), perceived complexity (Guan et al., 2022), or perceived stress (Han et al., 2022).

It is expected that computer vision methods, including the YOLO algorithm (Bochkovskiy et al., 2020) used in this study, will be increasingly important for understanding why drivers drive the way they do. Furthermore, the measurement techniques demonstrated here may be useful in the development of automated driving systems. Current AVs already exhibit cautious behavior when driving in the vicinity of humans in shared spaces (e.g., Tesla's full-self driving; AI DRIVR, 2022); accordingly, further consideration of perceived risk might enhance the overall driving experience.

5. Conclusion

This study examined the ability of a computer-vision algorithm to predict population-level perceived risk in still images of traffic scenes. It was found that basic features extracted from camera images combined with standard vehicle-state information in the form of vehicle speed allowed for a fairly strong prediction of perceived risk. These findings suggest that it may be possible to predict the perceived risk of a typical driver in an AV based on information obtained from AV sensors, which could be valuable in optimizing the ride experience in AVs without the need for complex driver-state monitoring technology. However, further research is needed to determine the predictive power of these features in actual AVs, in which speeds and accelerations of the ego-vehicle and other vehicles can be perceived, and in other types of driving environments.

CRediT authorship contribution statement

Joost de Winter: Conceptualization, Data curation, Formal analysis, Methodology, Validation, Visualization, Writing – original draft. **Jim Hoogmoed:** Investigation, Methodology, Resources. **Jork Stapel:** Conceptualization, Supervision. **Dimitra Dodou:** Writing – review & editing. **Pavlo Bazilinskyy:** Conceptualization, Data curation, Methodology, Resources, Software, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The questionnaire used in Appen, image stimuli, raw data, and a MATLAB script that reproduces the tables and figures are accessible at https://doi.org/10.4121/21952685. The code of the online experiment is available at https://github.com/bazilinskyy/risk-dash-crowdsourcing.

References

AI DRIVR. (2022). FSD Beta 10.12 pisses people off in SF. https://www.youtube.com/watch?v=yfuUI-5GBxs.

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, 21*, 347–353. https://doi.org/10.1080/15389588.2020.1762871

Bazilinskyy, P., Dodou, D., & De Winter, J. C. F. (2020). External Human-Machine Interfaces: Which of 729 colors is best for signaling 'Please (do not) cross'?. In *IEEE International Conference on Systems, Man and Cybernetics (SMC)* (pp. 3721–3728). https://doi.org/10.1109/SMC42975.2020.9282998

Bengler, K., Rettenmaier, M., Fritz, N., & Feierle, A. (2020). From HMI to HMIs: Towards an HMI framework for automated driving. *Information, 11*, 61. https://doi.org/10.3390/info11020061

Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). YOLOv4: Optimal speed and accuracy of object detection. arXiv 2020. https://arxiv.org/abs/2004.10934.

Brell, T., Philipsen, R., & Ziefle, M. (2019). sCARy! Risk perceptions in autonomous driving: The influence of experience on perceived benefits and barriers. *Risk Analysis, 39*, 342–357. https://doi.org/10.1111/risa.13190

Bustos, C., Elhaouij, N., Solé-Ribalta, A., Borge-Holthoefer, J., Lapedriza, A., & Picard, R. (2021). Predicting driver self-reported stress by analyzing the road scene. In *Proceedings of the 2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII)*. Japan: Nara. https://doi.org/10.1109/ACII52823.2021.9597438.

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

Cai, Q., Abdel-Aty, M., Zheng, O., & Wu, Y. (2022). Applying machine learning and Google street view to explore effects of drivers' visual environment on traffic safety. *Transportation Research Part C: Emerging Technologies, 135*, Article 103541. https://doi.org/10.1016/j.trc.2021.103541

Chapman, P. R., & Underwood, G. (1998). Visual search of driving situations: Danger and experience. *Perception, 27*, 951–964. https://doi.org/10.1068/p270951

Charlton, S. G., Mackie, H. W., Baas, P. H., Hay, K., Menezes, M., & Dixon, C. (2010). Using endemic road features to create self-explaining roads and reduce vehicle speeds. *Accident Analysis & Prevention, 42*, 1989–1998. https://doi.org/10.1016/j.aap.2010.06.006

Charlton, S. G., & Starkey, N. J. (2016). Risk in our midst: Centrelines, perceived risk, and speed choice. *Accident Analysis & Prevention, 95*, 192–201. https://doi.org/10.1016/j.aap.2016.07.019

Charlton, S. G., Starkey, N. J., Perrone, J. A., & Isler, R. B. (2014). What's the risk? A comparison of actual and perceived driving risk. *Transportation Research Part F: Traffic Psychology and Behaviour, 25*, 50–64. https://doi.org/10.1016/j.trf.2014.05.003

Chen, C., Zhao, X., Liu, H., Ren, G., & Liu, X. (2019). Influence of adverse weather on drivers' perceived risk during car following based on driving simulations. *Journal of Modern Transportation, 27*, 282–292. https://doi.org/10.1007/s40534-019-00197-4

De Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Webbrowser. *Behavior Research Methods, 47*, 1–12. https://doi.org/10.3758/s13428-014-0458-y

Detjen, H., Faltaous, S., Pfleging, B., Geisler, S., & Schneegass, S. (2021). How to increase automated vehicles' acceptance through in-vehicle interaction design: A review. *International Journal of Human-Computer Interaction, 37*, 308–330. https://doi.org/10.1080/10447318.2020.1860517

Elvik, R., Vadeby, A., Hels, T., & van Schagen, I. (2019). Updated estimates of the relationship between speed and road safety at the aggregate and individual levels. *Accident Analysis & Prevention, 123*, 114–122. https://doi.org/10.1016/j.aap.2018.11.014

Engwicht, D. (2012). Solving the safety paradox—When making things less safe makes them more safe. *Injury Prevention, 18*, A1. https://doi.org/10.1136/injuryprev-2012-040580i.2

Evans, T., Stuckey, R., & Macdonald, W. (2020). Young drivers' perceptions of risk and difficulty: Day versus night. *Accident Analysis & Prevention, 147*, Article 105753. https://doi.org/10.1016/j.aap.2020.105753

Fuller, R., McHugh, C., & Pender, S. (2008). Task difficulty and risk in the determination of driver behaviour. *European Review of Applied Psychology, 58*, 13–21. https://doi.org/10.1016/j.erap.2005.07.004

Gårder, P. E. (2004). The impact of speed and other variables on pedestrian safety in Maine. *Accident Analysis & Prevention, 36*, 533–542. https://doi.org/10.1016/S0001-4575(03)00059-9

Geiger, A., Lenz, P., Stiller, C., & Urtasun, R. (2013). Vision meets robotics: The KITTI dataset. *The International Journal of Robotics Research, 32*, 1231–1237. https://doi.org/10.1177/0278364913491297

Gibson, J. J., & Crooks, L. E. (1938). A theoretical field-analysis of automobile-driving. *The American Journal of Psychology, 51*, 453–471. https://doi.org/10.2307/1416145

Guan, F., Fang, Z., Wang, L., Zhang, X., Zhong, H., & Huang, H. (2022). Modelling people's perceived scene complexity of real-world environments using street-view panoramas and open geodata. *ISPRS Journal of Photogrammetry and Remote Sensing, 186*, 315–331. https://doi.org/10.1016/j.isprsjprs.2022.02.012

Guo, C., Kidono, K., Machida, T., Terashima, R., & Kojima, Y. (2017). Human-like behavior generation for intelligent vehicles in urban environment based on a hybrid potential map. *Proceedings of the 2017 IEEE Intelligent Vehicles Symposium (IV)*, Los Angeles, CA, 197–203. https://doi.org/10.1109/IVS.2017.7995720.

Han, X., Wang, L., Seo, S. H., He, J., & Jung, T. (2022). Measuring perceived psychological stress in urban built environments using Google Street View and deep learning. *Frontiers in Public Health, 10*, Article 891736. https://doi.org/10.3389/fpubh.2022.891736

Hancock, P. A., Kajaks, T., Caird, J. K., Chignell, M. H., Mizobuchi, S., Burns, P. C., … Vrkljan, B. H. (2020). Challenges to human drivers in increasingly automated vehicles. *Human Factors, 62*, 310–328. https://doi.org/10.1177/0018720819900402

Hardy, M., & Harvey, H. (2020). Artificial intelligence in diagnostic imaging: Impact on the radiography profession. *The British Journal of Radiology, 93*, 20190840. https://doi.org/10.1259/bjr.20190840

Hartwich, F., Beggiato, M., & Krems, J. F. (2018). Driving comfort, enjoyment and acceptance of automated driving – effects of drivers' age and driving style familiarity. *Ergonomics, 61*, 1017–1032. https://doi.org/10.1080/00140139.2018.1441448

He, X., Stapel, J., Wang, M., & Happee, R. (2022). Modelling perceived risk and trust in driving automation reacting to merging and braking vehicles. *Transportation Research Part F: Traffic Psychology and Behaviour, 86*, 178–195. https://doi.org/10.1016/j.trf.2022.02.016

Hecht, T., Feldhütter, A., Radlmayr, J., Nakano, Y., Miki, Y., Henle, C., & Bengler, K. (2019). A review of driver state monitoring systems in the context of automated driving. In S. Bagnara, R. Tartaglia, S. Albolino, T. Alexander, & Y. Fujita (Eds.), *Proceedings of the 20th Congress of the International Ergonomics Association* (pp. 398–408). Cham: Springer. https://doi.org/10.1007/978-3-319-96074-6_43.

Heikoop, D. D., De Winter, J. C. F., Van Arem, B., & Stanton, N. A. (2016). Psychological constructs in driving automation: A consensus model and critical comment on construct proliferation. *Theoretical Issues in Ergonomics Science, 17*, 284–303. https://doi.org/10.1080/1463922X.2015.1101507

Heikoop, D. D., De Winter, J. C. F., Van Arem, B., & Stanton, N. A. (2017). Effects of platooning on signal-detection performance, workload, and stress: A driving simulator study. *Applied Ergonomics, 60*, 116–127. https://doi.org/10.1016/j.apergo.2016.10.016

Hulse, L. M., Xie, H., & Galea, E. R. (2018). Perceptions of autonomous vehicles: Relationships with road users, risk, gender and age. *Safety Science, 102*, 1–13. https://doi.org/10.1016/j.ssci.2017.10.001

Kinnear, N., Kelly, S. W., Stradling, S., & Thomson, J. (2013). Understanding how drivers learn to anticipate risk on the road: A laboratory experiment of affective anticipation of road hazards. *Accident Analysis & Prevention, 50*, 1025–1033. https://doi.org/10.1016/j.aap.2012.08.008

Kinnear, N., Stradling, S., & McVey, C. (2008). Do we really drive by the seat of our pants?. In L. Dorn (Ed.), *Driver behaviour and training* (Vol. 3, pp. 349–365) Aldershot, UK: Ashgate.

Kolekar, S., De Winter, J., & Abbink, D. (2020). Which parts of the road guide obstacle avoidance? Quantifying the driver's risk field. *Applied Ergonomics, 89*, Article 103196. https://doi.org/10.1016/j.apergo.2020.103196

Kolekar, S., Petermeijer, B., Boer, E., De Winter, J., & Abbink, D. (2021). A risk field-based metric correlates with driver's perceived risk in manual and automated driving: A test-track study. *Transportation Research Part C: Emerging Technologies, 133*, Article 103428. https://doi.org/10.1016/j.trc.2021.103428

Kondoh, T., Yamamura, T., Kitazaki, S., Kuge, N., & Boer, E. R. (2008). Identification of visual cues and quantification of drivers' perception of proximity risk to the lead vehicle in car-following situations. *Journal of Mechanical Systems for Transportation and Logistics, 1*, 170–180. https://doi.org/10.1299/jmtl.1.170

Kooijman, B. (2021). *The identification of factors affecting drivers' perceived risk in pedestrian-vehicle interaction: A crowdsourcing study*. Delft University of Technology. MSc thesis.

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences, 110*, 5802–5805. https://doi.org/10.1073/pnas.1218772110

Kovácsová, N., De Winter, J. C. F., & Hagenzieker, M. P. (2019). What will the car driver do? A video-based questionnaire study on cyclists' anticipation during safety-critical situations. *Journal of Safety Research, 69*, 11–21. https://doi.org/10.1016/j.jsr.2019.01.002

Kwon, J.-H., & Cho, G.-H. (2020). An examination of the intersection environment associated with perceived crash risk among school-aged children: Using street-level imagery and computer vision. *Accident Analysis & Prevention, 146*, Article 105716. https://doi.org/10.1016/j.aap.2020.105716

Lee, Y. M., Miller, K., Crundall, D., & Sheppard, E. (2020). Cross-cultural effects on detecting multiple sources of driving hazard: Evidence from the deceleration detection flicker test. *Transportation Research Part F: Traffic Psychology and Behaviour, 69*, 222–234. https://doi.org/10.1016/j.trf.2020.01.013

Lefèvre, S., Carvalho, A., Gao, Y., Tseng, H. E., & Borrelli, F. (2015). Driver models for personalised driving assistance. *Vehicle System Dynamics, 53*, 1705–1720. https://doi.org/10.1080/00423114.2015.1062899

Lewis-Evans, B., De Waard, D., & Brookhuis, K. A. (2010). That's close enough—A threshold effect of time headway on the experience of risk, task difficulty, effort, and comfort. *Accident Analysis & Prevention, 42*, 1926–1933. https://doi.org/10.1016/j.aap.2010.05.014

Lewis-Evans, B., & Rothengatter, T. (2009). Task difficulty, risk, effort and comfort in a simulated driving task—Implications for Risk Allostasis Theory. *Accident Analysis & Prevention, 41*, 1053–1063. https://doi.org/10.1016/j.aap.2009.06.011

Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft COCO: Common Objects in Context. Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 740–755. https://doi.org/10.1007/978-3-319-10602-1_48.

Liu, Y., & Salvendy, G. (2009). Effects of measurement errors on psychometric measurements in ergonomics studies: Implications for correlations, ANOVA, linear regression, factor analysis, and linear discriminant analysis. *Ergonomics, 52*, 499–511. https://doi.org/10.1080/00140130802392999

Lu, Z., & De Winter, J. C. F. (2015). A review and framework of control authority transitions in automated driving. *Proceedings of the 6th International Conference on Applied Human Factors and Ergonomics (AHFE)*, Las Vegas, NV, 2510–2517. https://doi.org/10.1016/j.promfg.2015.07.513.

McKenna, F. P., & Crick, J. (1997). *Developments in hazard perception; prepared for road safety division of DETR* (TRL Report 297). Crowthorne, Berkshire, UK: Transport Research Laboratory TRL.

Megías, A., Cándido, A., Maldonado, A., & Catena, A. (2018). Neural correlates of risk perception as a function of risk level: An approach to the study of risk through a daily life task. *Neuropsychologia, 119*, 464–473. https://doi.org/10.1016/j.neuropsychologia.2018.09.012

Melman, T., Abbink, D. A., Van Paassen, M. M., Boer, E. R., & De Winter, J. C. F. (2018). What determines drivers' speed? A replication of three behavioural adaptation experiments in a single driving simulator study. *Ergonomics, 61*, 966–987. https://doi.org/10.1080/00140139.2018.1426790

Merat, N., Seppelt, B., Louw, T., Engström, J., Lee, J. D., Johansson, E., Green, C. A., Katazaki, S., Monk, C., Itoh, M., McGehee, D., Sunda, T., Unoura, K., Victor, T., Schieben, A., & Keinath, A. (2019). The "Out-of-the-Loop" concept in automated driving: Proposed definition, measures and implications. *Cognition, Technology & Work, 21*, 87–98. https://doi.org/10.1007/s10111-018-0525-8

Morales-Alvarez, W., Sipele, O., Léberon, R., Tadjine, H. H., & Olaverri-Monreal, C. (2020). Automated driving: A literature review of the take over request in conditional automation. *Electronics, 9*, 2087. https://doi.org/10.3390/electronics9122087

Moran, C., Bennett, J. M., & Prabhakharan, P. (2019). Road user hazard perception tests: A systematic review of current methodologies. *Accident Analysis & Prevention, 129*, 309–333. https://doi.org/10.1016/j.aap.2019.05.021

Näätänen, R., & Summala, H. (1974). A model for the role of motivational factors in drivers' decision-making. *Accident Analysis & Prevention, 6*, 243–261. https://doi.org/10.1016/0001-4575(74)90003-7

Nagle, F., & Lavie, N. (2020). Predicting human complexity perception of real-world scenes. *Royal Society Open Science, 7*, Article 191487. https://doi.org/10.1098/rsos.191487

Niermann, D., Trende, A., Ihme, K., Drewitz, U., Hollander, C., & Hartwich, F. (2021). An integrated model for user state detection of subjective discomfort in autonomous vehicles. *Vehicles, 3*, 764–777. https://doi.org/10.3390/vehicles3040045

Nordfjærn, T., Jørgensen, S., & Rundmo, T. (2011). A cross-cultural comparison of road traffic risk perceptions, attitudes towards traffic safety and driver behaviour. *Journal of Risk Research, 14*, 657–684. https://doi.org/10.1080/13669877.2010.547259

Nunnally, J. C. (1967). *Psychometric theory*. New York: McGraw-Hill Book Company.

Ohn-Bar, E., & Trivedi, M. M. (2017). Are all objects equal? Deep spatio-temporal importance prediction in driving videos. *Pattern Recognition, 64*, 425–436. https://doi.org/10.1016/j.patcog.2016.08.029

Pascale, M. T., Rodwell, D., Coughlan, P., Kaye, S.-A., Demmel, S., Dehkordi, S. G., Bond, A., Lewis, I., Rakotonirainy, A., & Glaser, S. (2021). Passengers' acceptance and perceptions of risk while riding in an automated vehicle on open, public roads. *Transportation Research Part F: Traffic Psychology and Behaviour, 83*, 274–290. https://doi.org/10.1016/j.trf.2021.10.012

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

Radhakrishnan, V., Merat, N., Louw, T., Lenné, M. G., Romano, R., Paschalidis, E., Hajiseyedjavadi, F., Wei, C., & Boer, E. R. (2020). Measuring drivers' physiological response to different vehicle controllers in highly automated driving (HAD): Opportunities for establishing real-time values of driver discomfort. *Information, 11*, 390. https://doi.org/10.3390/info11080390

Rashed, H., Ramzy, M., Vaquero, V., El Sallab, A., Sistu, G., & Yogamani, S. (2019). FuseMODNet: Real-time camera and LiDAR based moving object detection for robust low-light autonomous driving. *Proceedings of IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, Seoul, Korea. https://doi.org/10.1109/ICCVW.2019.00293.

Rasouli, A., Kotseruba, I., & Tsotsos, J. K. (2018). Understanding pedestrian behavior in complex traffic scenes. *IEEE Transactions on Intelligent Vehicles, 3*, 61–70. https://doi.org/10.1109/TIV.2017.2788193

Rossner, P., & Bullinger, A. C. (2020). How do you want to be driven? investigation of different highly-automated driving styles on a highway scenario. In N. Stanton (Ed.), *Advances in Human Factors of Transportation (AHFE)* (pp. 36–43). Cham: Springer. https://doi.org/10.1007/978-3-030-20503-4_4.

Rudenko, A., Palmieri, L., Herman, M., Kitani, K. M., Gavrila, D. M., & Arras, K. O. (2020). Human motion trajectory prediction: A survey. *The International Journal of Robotics Research, 39*, 895–935. https://doi.org/10.1177/0278364920917446

Rushton, J. P., Brainerd, C. J., & Pressley, M. (1983). Behavioral development and construct validity: The principle of aggregation. *Psychological Bulletin, 94*, 18–38. https://doi.org/10.1037/0033-2909.94.1.18

Saffarian, M., Happee, R., & De Winter, J. C. F. (2012). Why do drivers maintain short headways in fog? A driving-simulator study evaluating feeling of risk and lateral control during automated and manual car following. *Ergonomics, 55*, 971–985. https://doi.org/10.1080/00140139.2012.691993

sbairagy-MW. (2021). Pretrained-yolo-v4. https://github.com/matlab-deep-learning/pretrained-yolo-v4.

Schömig, N., Hargutt, V., Neukum, A., Petermann-Stock, I., & Othersen, I. (2015). The interaction between highly automated driving and the development of drowsiness. *Procedia Manufacturing, 3*, 6652–6659. https://doi.org/10.1016/j.promfg.2015.11.005

Siebert, F. W., Oehl, M., & Pfister, H.-R. (2014). The influence of time headway on subjective driver states in adaptive cruise control. *Transportation Research Part F: Traffic Psychology and Behaviour, 25*, 65–73. https://doi.org/10.1016/j.trf.2014.05.005

Stapel, J., Gentner, A., & Happee, R. (2022). On-road trust and perceived risk in Level 2 automation. *Transportation Research Part F: Traffic Psychology and Behaviour, 89*, 355–370. https://doi.org/10.1016/j.trf.2022.07.008

Tan, H., Lu, G., & Liu, M. (2022). Risk field model of driving and its application in modeling car-following behavior. *IEEE Transactions on Intelligent Transportation Systems*. https://doi.org/10.1109/TITS.2021.3105518

Taylor, D. H. (1964). Drivers' galvanic skin response and the risk of accident. *Ergonomics, 7*, 439–451. https://doi.org/10.1080/00140136408930761

Van der Horst, R., & Kaptein, N. (1998). Self-explaining roads. *Proceedings of the 11th ICTCT Workshop*, Budapest, Hungary, 15–32. https://www.ictct.net/wp-content/uploads/11-Budapest-1998/ictct_document_nr_307_2.pdf.

Van Paassen, M. M., Boink, R. P., Abbink, D. A., Mulder, M., & Mulder, M. (2015). Haptic guidance, interaction between the guidance model and tuning. *Proceedings of the 18th International Symposium on Aviation Psychology*, Dayton, OH, 410–415. https://corescholar.libraries.wright.edu/isap_2015/37/.

Ventsislavova, P., Crundall, D., Baguley, T., Castro, C., Gugliotta, A., Garcia-Fernandez, P., Zhang, W., Ba, Y., & Li, Q. (2019). A comparison of hazard perception and hazard prediction tests across China, Spain and the UK. *Accident Analysis & Prevention, 122*, 268–286. https://doi.org/10.1016/j.aap.2018.10.010

Vlakveld, W. P. (2014). A comparative study of two desktop hazard perception tasks suitable for mass testing in which scores are not based on response latencies. *Transportation Research Part F: Traffic Psychology and Behaviour, 22*, 218–231. https://doi.org/10.1016/j.trf.2013.12.013

Weaver, B. W., & DeLucia, P. R. (2020). A systematic review and meta-analysis of takeover performance during conditionally automated driving. *Human Factors*. https://doi.org/10.1177/0018720820976476

Wilde, G. J. S. (1982). The theory of risk homeostasis: Implications for safety and health. *Risk Analysis, 2*, 209–225. https://doi.org/10.1111/j.1539-6924.1982.tb01384.x

Wilde, G. J. S. (2013). Homeostasis drives behavioural adaptation. In C. M. Rudin-Brown, & S. L. Jamson (Eds.), *Behavioural adaptation and road safety: Theory, evidence and action* (pp. 61–86). Boca Raton, FL: CRC Press.

Yurtsever, E., Liu, Y., Lambert, J., Miyajima, C., Takeuchi, E., Takeda, K., & Hansen, J. H. L. (2019). Risky action recognition in lane change video clips using deep spatiotemporal networks with segmentation mask transfer. *Proceedings of the 2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, Auckland, New Zealand, 3100–3107. https://doi.org/10.1109/ITSC.2019.8917362.

Zhang, B., Wilschut, E. S., Willemsen, D. M. C., & Martens, M. H. (2019). Transitions to manual control from highly automated driving in non-critical truck platooning scenarios. *Transportation Research Part F: Traffic Psychology and Behaviour, 64*, 84–97. https://doi.org/10.1016/j.trf.2019.04.006

Zhao, X., He, R., & Wang, J. (2020). How do drivers respond to driving risk during car-following? Risk-response driver model and its application in human-like longitudinal control. *Accident Analysis & Prevention, 148*, Article 105783. https://doi.org/10.1016/j.aap.2020.105783