Predicting Perceived Risk of Traffic Scenes using Computer Vision

August 2022

Joost de Winter*, Jim Hoogmoed, Jork Stapel, Dimitra Dodou, Pavlo Bazilinskyy

Department of Cognitive Robotics, Faculty of Mechanical Engineering, Delft University of Technology, The Netherlands

*Corresponding author, j.c.f.dewinter@tudelft.nl

Abstract

Perceived (i.e., subjective) risk is a key construct in traffic psychology and, more recently, in the area of automated driving. If the automated vehicle would be able to estimate the risk perceived by its occupants, this would allow the automated vehicle to adjust its behavior accordingly. This paper examines whether perceived risk in images of traffic scenes is predictable from computer-vision features that may also be used by AVs. A total of 1378 participants rated the perceived risk for a random 100 of 210 images from a driver's perspective obtained from the KITTI dataset. The population-level perceived risk was found to be statistically reliable (r = 0.98 split-half reliability). Linear regression showed that perceived risk was predictable (r = 0.62) from two features obtained with the YOLOv4 computer-vision algorithm: the number of persons in the scene and the mean size of the bounding boxes surrounding other road users. In a second regression model, the ego-vehicle's speed was added as an independent variable, which substantially increased the strength of the prediction to r = 0.75. Paradoxically, the sign of the speed prediction was negative; that is, a higher vehicle speed implied a lower perceived risk, a finding that resonates with the principle of self-explaining roads. It is concluded that objective features can contribute to an accurate prediction of population subjective risk in a way that outperforms the ratings provided by individual participants (mean r = 0.41). The present findings may be relevant to automated-vehicle development and to the modeling of psychological constructs in the field of traffic psychology.

Introduction

Much attention is currently being paid to the development of automated vehicles (AVs). In addition to the technical challenges involved in creating AVs, there are various human factors challenges as well (e.g., Hancock et al., 2020; Merat et al., 2019). It seems that most human factors research so far is concerned with measuring how drivers respond when having to take over control from the AV (Lu & De Winter, 2015; Morales-Alvarez et al., 2020; Weaver & DeLucia, 2020; Zhang et al., 2019) as well as human-machine interface design (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 for several hours in a simulator. Participants reported a reduction in energetic arousal and concentration compared to before the experiment but were found to remain able to detect salient events in the environment. Other researchers quantified driver drowsiness during automated driving from measurements of eyelid closure and posture (e.g., Hecht et al., 2019; Schömig et al., 2015). Several studies have used a handset control unit to continuously poll drivers' discomfort with the automated vehicle'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).

Apart from driver state variables such as drowsiness and comfort, perceived (i.e., subjective) risk may be an important variable in automated driving. In manual driving, perceived risk has been regarded as a key construct that mediates the driver's actions (Charlton & Starkey, 2016; Näätänen & Summala, 1974; Wilde, 1982, 2013). It would be relevant to determine the perceived risk of drivers in the AV for several reasons. First, if the AV can determine perceived risk, then the AV can determine whether or not it is safe to hand back control to the human driver (Cabrall et al., 2020; Kolekar et al., 2021). A second reason for determining perceived risk is that AV controllers may consider this variable to optimize the driving experience. That is, the AV could slow down when perceived risk is anticipated to be high, and the consideration of risk in AV controllers may help ensure that the AV drives in a manner that is accepted (Guo et al., 2017; Kolekar et al., 2020; Rossner & Bullinger, 2020; Tan et al., 2022; Zhao et al., 2020).

Various studies have examined how perceived risk relates to the driving conditions. A common approach in hazard perception studies is to let observers watch video clips and press a response key when they detect a hazard (Chapman & Underwood, 1998; McKenna & Crick, 1997; Vlakveld, 2014; for a review, see Moran et al., 2019). Using the same approach, Bazilinskyy et al. (2020) measured perceived risk of a large number of participants in different countries by letting them press a key on the keyboard while viewing dashcam video clips. Their results showed that perceived risk varied considerably during the drive, with mostly uneventful scenarios followed by bursts in risk, for example, when approaching pedestrians or other cars. Other studies showed that perceived risk is affected by visibility, such as night versus daytime driving (Evans et al., 2020) and driving in fog versus clear weather (Chen et al., 2019; Saffarian et al., 2012). Perceived risk is also affected by 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), among others.

The above studies are informative about factors that contribute to perceived risk but only focus on the *measurement* of risk, often in pre-programmed scenarios in a driving simulator. It may be of considerable value to be able to *predict* perceived risk solely based on vehicle state and camera images, that is, sensors available in AVs. To date, several studies have used this approach. 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 video frames overlaid with these masks were subsequently fed into a Long Short-Term Memory (LSTM) architecture. A similar approach was used by Ping et al. (2018). They let human

annotators indicate the risk of individual frames of dashcam videos and 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 from their verbal utterances and from whether drivers 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) exposed participants to short dashcam video clips and asked drivers to report the perceived effort to reclaim control from the AV. They showed 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 the image characteristics. More specifically, 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. Additionally, the method of Ping et al. (2018) was partially subjective since 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, it appears that the above works treat risk measurement as a typical labeling/annotation task in computer vision research, where a small number of participants rate 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 appeared to give little consideration to psychometric principles, such as whether the risk annotation is statistically reliable, where reliability can be seen as a necessary condition for validity (Nunnally, 1967). More specifically, there are concerns within the field of psychology that 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).

In this study, our objective was to assess whether perceived risk, determined by a large number of human raters, could be predicted from basic features obtained using computer vision, as would also be expected to be available in an AV. In the current study, we measured participants' perceived risk for diverse photos of traffic scenes. Photos do not reveal information about speed but carry a number of advantages compared to videos. In videos, it may be ambiguous whether risk perception should be considered excluding or including the driver's response. For example, if, in a video, a driver approaches a busy traffic situation and brakes, perceived risk could be seen as low (because the driver appropriately slowed down) or as high (because a busy traffic situation is ahead). It can be argued that the use of images constitutes a pure stimulus-driven way of assessing perceived risk.

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. 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, and the time intervals between images of a drive were 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 x 375 pixels.

A crowdsourcing approach was used to gather risk ratings per image. We allowed 2000 contributors from all countries to participate. The research 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 asked to leave the questionnaire by clicking on a link that opened a webpage with the experiment, which was created using jsPsych (De Leeuw, 2015). It presented participants 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. The reason for not presenting all 210 images was to prevent potential boredom during the study. Below each image, it was mentioned: "*As a driver, how risky would you judge this situation (0 = no risk, 10 = extreme risk)?*" An example is shown in Figure 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 in the Appen page. A payment of USD 0.25 was provided for the completion of the study.

Computer vision was used to detect objects in the images. The KITTI dataset contains manually labeled bounding boxes, but 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 Figure 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.

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

Figure 2. Results of YOLOV4 on 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.

Results

Data filtering

The 2000 respondents took part between 23 November and 17 December 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 not to have read the instructions, participants who completed the study in less than 5

min, and participants who 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, 622 of the 2000 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).

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, it was examined whether the PPR was statistically reliable. Figure 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 risk across the rated images with 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). Figure 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 (32.7%, n = 591), USA (37.8%, n = 95), Russia (27.1%, n = 68), Ukraine (24.2%, n = 64), and India (34.5%, n = 57). Figure 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 Figures 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 = 0.018), barely higher than the r = 0.82 association depicted in Figure 5.

In other words, it can be concluded that although there may be national differences in perceived risk (see also Bazilinskyy et al., 2020; Lee et al., 2020; Nordfjærn et al., 2011; Ventsislavova et al., 2019), the relative differences in PPR values between images seems highly invariant to nationality, as well as age, gender, and driving experience.
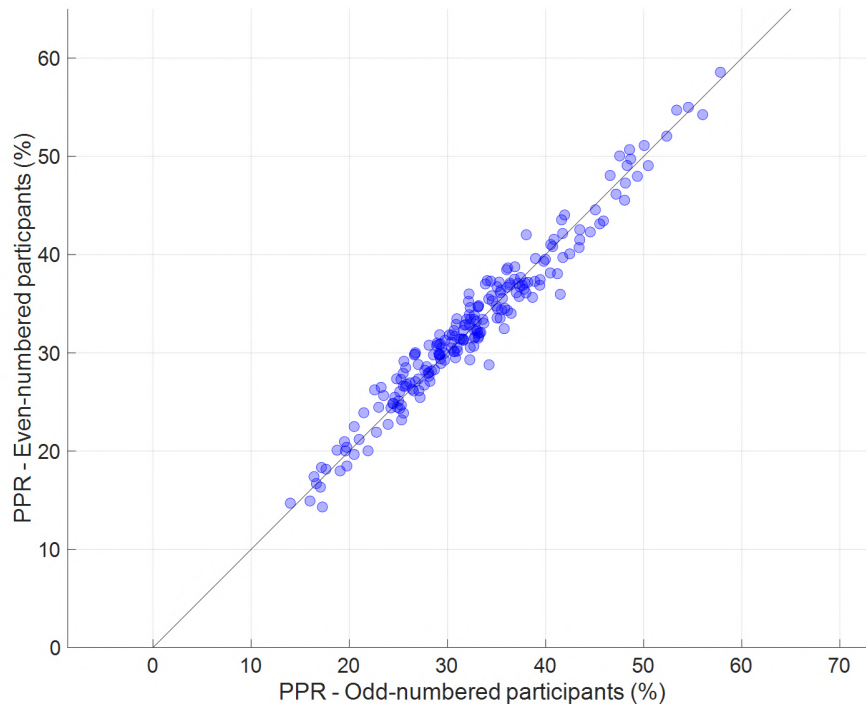
Figure 3. Mean perceived risk per image as reported by even-numbered participants (*n* = 689) versus mean perceived risk as reported by odd-numbered participants (*n* = 689). Each marker represents an image and is based on an average of 326 responses.
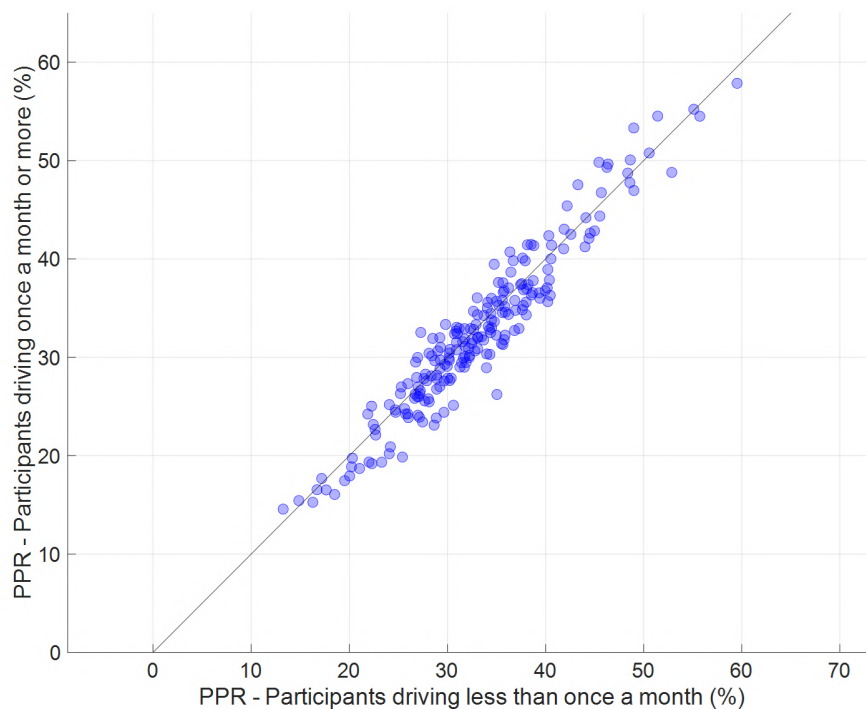

Figure 4. Mean perceived risk per image for participants who reported driving less than a month (*n* = 281) versus participants who reported driving once a month or more frequently (*n* = 1079). Each marker represents an image and is based on an average of 133 and 511 responses for the former and latter groups of participants.
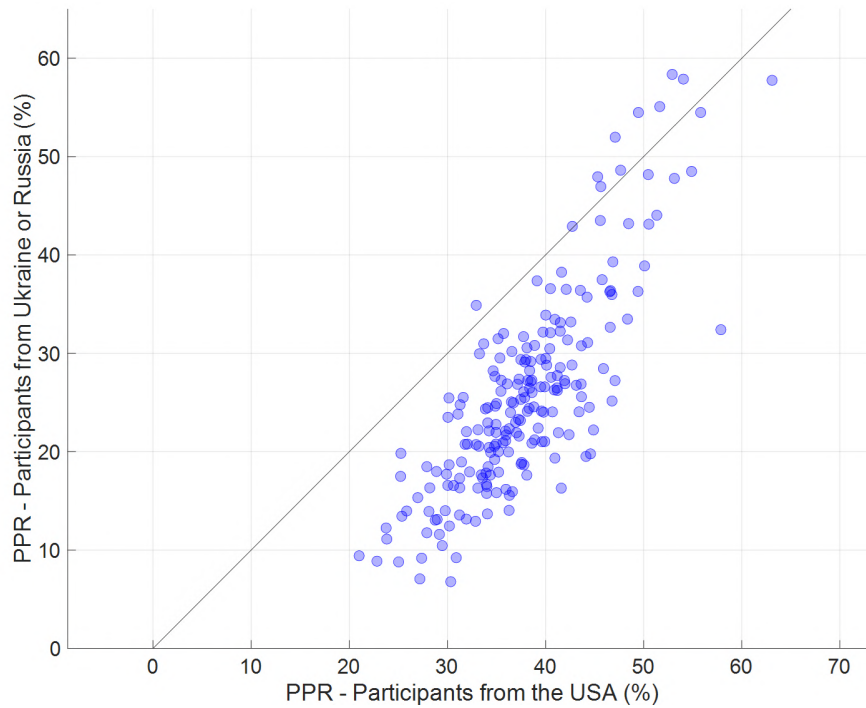
Figure 5. Mean perceived risk per image for participants from Ukraine and Russia combined (*n* = 132) versus participants from the USA (*n* = 95). Each marker represents an image and is based on an average of 63 and 45 responses for the former and latter groups of participants.

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).

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

Two features were extracted from the YOLOv4 bounding box results. These were the following:

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

    Psychologically, pedestrians imply an unpredictable environment. Figure 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)*:

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

    , where i is the instance number, n is the number of instances for Persons, Bicycles, Cars, Motorbikes, Buses, Trains, or Truck, 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. Figure 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.

Figure 6 provides eight images, sorted from lowest to highest perceived risk in eight octiles, together with the scores of the two features (Persons and Area), as well as vehicle speed obtained from the KITTI dataset. It can be seen that the lowest PPR occurred for a completely straight and empty road. The highest PPR was found for an image with a car that was frontally approaching, as represented by a relatively large Area score.

- PPR rank: 1st (lowest risk)
- PPR: 14.3%
- Number of persons: 0
- Area: 0 px
- Speed: 21.6 m/s

- PPR rank: 30th
- PPR: 24.7%
- Number of persons: 0
- Area: 0 px
- Speed: 12.9 m/s

- PPR rank: 60th
- PPR: 28.3%
- Number of persons: 0
- Area: 29 px
- Speed: 12.5 m/s

- PPR rank: 90th
- PPR: 31.0%
- Number of persons: 0
- Area: 0 px
- Speed: 7.7 m/s

- PPR rank: 120th
- PPR: 33.0%
- Number of persons: 0
- Area: 96 px
- Speed: 9.4 m/s

- PPR rank: 150th
- PPR: 35.9%
- Number of persons: 1
- Area: 104 px
- Speed: 9.1 m/s

- PPR rank: 180th
- PPR: 40.7%
- Number of persons: 1
- Area: 55 px
- Speed: 2.7 m/s

- PPR rank: 210th (highest risk)
- PPR: 58.2%
- Number of persons: 0
- Area: 124 px
- Speed: 0.6 m/s

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

Predicting Population Perceived Risk

The correlation matrix among the measures shows that the two computer-vision features correlated substantially with PPR (Table 2). A third measure, vehicle speed, showed a strong *negative* correlation with PPR. Furthermore, it can be seen that the Area measure was generally smaller when vehicle speed was higher. A possible explanation is that drivers adopt larger margins to other vehicles on higher-speed roads. Table 2 also shows correlations with the median time it took participants to enter their response, which reveals that images that were perceived to be less risky were rated more quickly.

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

Results from a linear regression analysis using the number of persons and Area as predictors of perceived risk are provided in Table 3. The predictive correlation of the weighted mean of the number of persons and Area equalled $r = 0.62$. Vehicle speed is a variable that was not accessible to participants in the crowdsourcing study, but it is a variable readily available in manual and automated vehicles. Table 4 shows the results of the regression analysis, now with the addition of vehicle speed. The predictive correlation of the regression model is $r = 0.75$ and is illustrated in Figure 7. An example of an image where risk was misestimated is shown in Figure 8. Here, participants gave higher risk ratings than the model.

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

	Unstandardized B	Standardized	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

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

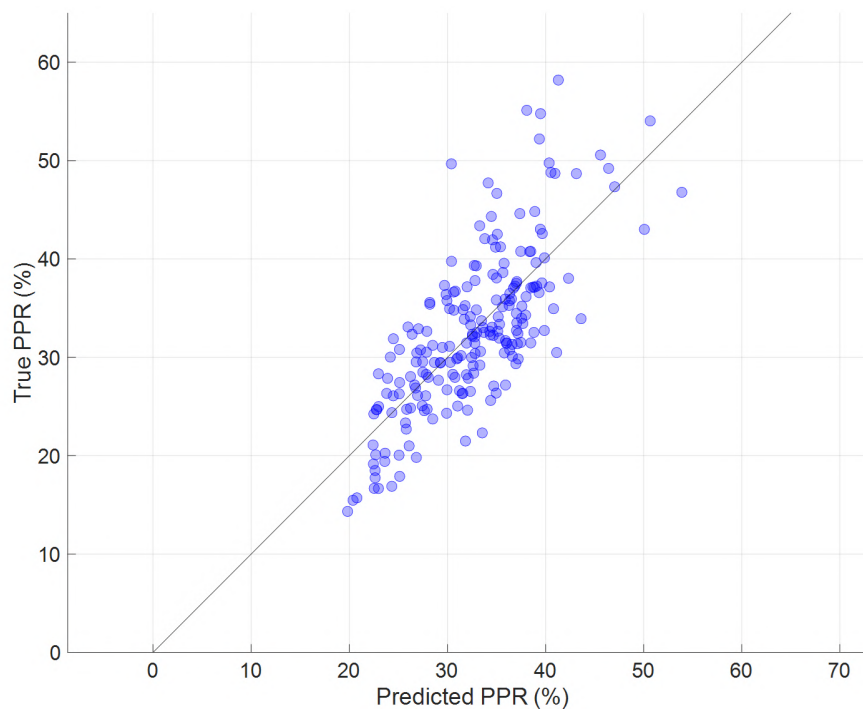|	Unstandardized B	Standardized	t	p
Intercept	34.90			
Number of persons	2.32	0.27	5.73	< 0.001
Area (px)	0.05	0.33	6.52	< 0.001
Speed (m/s)	-0.70	-0.46	-9.11	< 0.001

Figure 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.

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

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 Figure 9. The mean correlation coefficient is 0.41, and the 1th, 5th, 25th, 50th, 75th, 95th, and 99th percentiles are -0.24, -0.02, 0.29, 0.45, 0.57, 0.69, and 0.75. In other words, a participant's risk ratings for the 100 rated images correlated on average 0.41 with the PPR values for the same images. These results also show that computer vision has the potential to outperform human raters. More specifically, the regression model shown in Table 3 yielded a predictive *r* that corresponded to the 84.9th percentile, while the regression model shown in Table 4 performed at the 98.6th percentile.
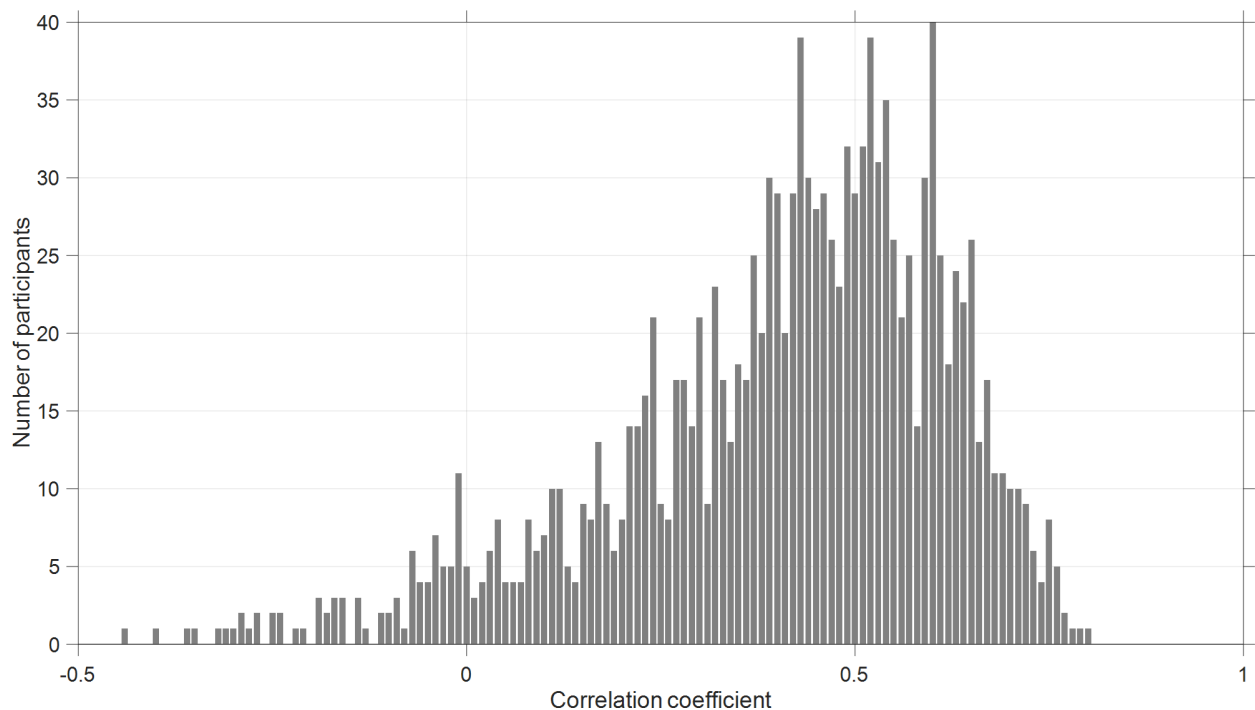

Figure 9. Distribution of observed correlation coefficients between participants' (*n* = 1378) perceived risk of the 100 images they rated and the population perceived risk (PPR) of the same images.

Discussion

Automated vehicles (AV) are developed to navigate safely through traffic but are not necessarily attuned to how the driver inside the vehicle feels. In particular, it may be valuable if the driver's perceived risk could be predicted from the AV's sensor information alone. Information on the estimated perceived risk would allow for optimizing the ride in the AV, without having to rely on a complex and sensitive apparatus such as driver-state monitoring.

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

The current findings resonate with research that showed 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 persons' Facebook likes. Also using regression analysis, Nagle and Lavie (2020) found that participants' ratings of 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 for traffic psychology. While perceived risk has previously been measured using 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; 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 deemed psychologically plausible. That is, the number of persons as a risk indicator aligns with the fact that driver-pedestrians interaction is a long-known issue in traffic safety, which is, in part, 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 notion 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 because they were a strong predictor of perceived risk in the KITTI dataset, which featured shared spaces (see Figure 2, upper panel). It can be imagined that the number of persons will 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.

A limitation of this study is that it used still images, for reasons outlined in the introduction. The use of images is further justified 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 et al. (2020) found that an unexpected stationary vehicle on the highway caused a burst in 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 Figure 8. In this case, humans rated the risk level to be high, but the predicted risk was low because other road users were still quite far away. A plausible explanation is the railroad that crossed the main road. Presumably, the driver in the video had already assessed that the situation was safe to cross, but the small field-of-view dashcam screenshots may have failed to give that impression to participants.

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 (or other types of self-ratings, as well as physiological measurements, etc.) are known to be statistically unreliable (see also the weak participant-population correlation of perceived risk of only 0.41 in this study). This makes it practically hard 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 used (e.g., Lefèvre et al., 2015; Van Paassen et al., 2015).

There has also been debate about whether risk measurements in traffic psychology should be concerned with 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). Also, it may be questioned whether perceived risk is sufficiently distinct from other often-used and strongly correlated constructs such as effort, comfort, trust, etc. (He et al., 2022; Siebert et al., 2014), an issue that relates to the problem of construct proliferation (Heikoop et al., 2016). It is our contention that similar results would have been obtained if, instead of perceived risk, we had used another rating scale, such as perceived effort (Cabrall et al., 2020).

It is possible that the use of images was a contributor to the *negative* correlation between ego-vehicle speed and perceived risk since participants may believe that risk is high in

situations in which the ego-vehicle was driving slowly. For example, the highest-risk image (Figure 6, bottom) would likely not be regarded as high-risk if the participant could see from a video that the ego-vehicle was almost standing still. 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 found a strong correlation between perceived risk and ego-vehicle speed as well (r = -0.45, r = -0.52; Kooijman, 2021). A plausible explanation is that lower-risk environments such as straight roads with clear priority rules enable faster driving (Kooijman, 2021). Another illustration is provided by the 90th-risky image in Figure 3: Even though there are zero detected road users, perceived risk is moderately elevated compared to a straight and empty road (lowest risk image). The elevated risk may be explained by the upcoming curve, for which the ego-driver likely slowed down (the ego-vehicle speed was only 7.7 m/s). In other words, a slower speed is indicative of risk because the ego-driver decided to drive more slowly. It is plausible that the inclusion of 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 encompasses the idea that perceptual features of a 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), while across situations a higher speed is associated with reduced risk. It is this paradox that is exploited in the design of shared spaces, following the principles of Monderman (Engwicht, 2012).

Clearly, the present study provided a basic showcase of the opportunities of computer vision by only considering the number of persons and the proximity of objects, as well as vehicle speed, as key predictors. Other research shows that various other factors contribute to perceived risk, such as road maintenance, infrastructural issues, brake lights, but also overtaking maneuvers or lateral maneuvers of other vehicles (Bazilinskyy et al., 2020; Yurtsever et al., 2019). It can be imagined that computer vision methods or the use of neural networks on image pixels directly (e.g., Nagle & Lavie, 2020) may contribute to a stronger prediction of perceived risk. The present study was concerned with perceived risk of a traffic situation from the driver's perspective. Other opportunities are to use a more location-specific approach by examining the relationship between (panoramic) images of road environments (such as obtained using Google Street View) and objective risk, that is crash likelihood (Cai et al., 2022), or to use such images to examine how road and city design contributes 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 of increasing relevance to obtaining an understanding of why drivers drive the way they do. Furthermore, the measurement techniques demonstrated in this study may be of benefit to the development of automated driving systems. The most recent automated vehicles already demonstrate cautious driving around persons in shared spaces (see Tesla's full-self driving; AI DRIVR, 2022), and further consideration of perceived risk may benefit the overall driving experience.

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

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. *Proceedings of the 2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII)*, Nara, Japan. 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*, 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*, 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*, 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., Feng, J., Fernie, G. R., Lavallière, M., Noy, I. Y., Redelmeier, D. A., & 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*, 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*, 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* (MSc thesis). Delft University of Technology. http://resolver.tudelft.nl/uuid:f3456522-18ec-4c91-8207-c4770b46322f

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*, 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. P*roceedings 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*, 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

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*, 105783. https://doi.org/10.1016/j.aap.2020.105783