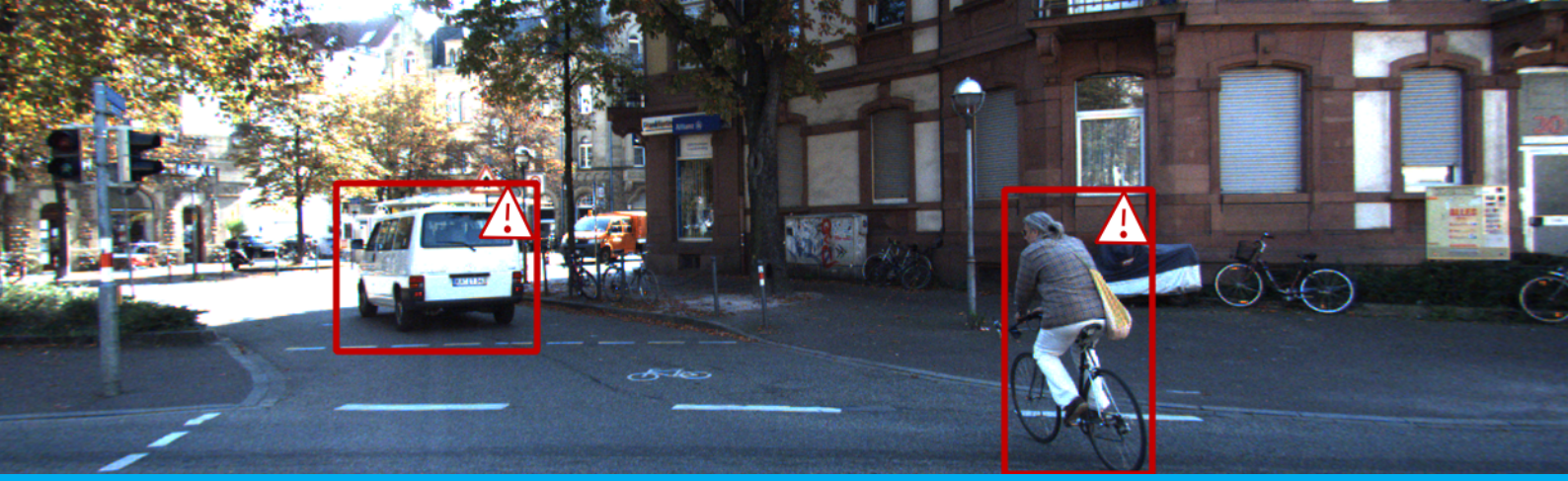

Technische Universiteit Delft

Human driver risk perception model

Fundamental threat parameters and what makes a driving situation risky

Jim Hoogmoed

TU Delft

Human driver risk perception model

Fundamental threat parameters and what makes a driving situation risky

by

J. (Jim) Hoogmoed

to obtain the degree of
Master of Science
in Mechanical Engineering
at the Delft University of Technology

Student Number:	4208056	
Academic staff:	Dr. Ir. J. (Joost) De Winter	Supervisor
	Dr. Ir. J.C.J. (Jork) Stapel	Daily supervisor
	Dr. Ir. P. (Pavlo) Bazilinskyy	Daily supervisor

Published on: August 31, 2021

TU Delft

Abstract

The level of automation in vehicles is growing. But until all vehicles are completely automated, there will be a transition period where automated vehicles and human drivers coexist. Because these road users will coexist, it is necessary that automated vehicles understand human drivers and vice versa. This study aims to create a model that predicts human risk perception in different driving scenarios, to provide an understanding of the fundamental features of human threat perception while driving.

The model created is a multi-criteria decision-making process that uses KITTI Vision Benchmark data as an input. This model is tested against the data gathered by an online survey, where 1918 participants answered the question: "How high is the risk on a scale from 0-10?" for 100 situations, chosen from the KITTI Vision Benchmark data. The survey response data is then compared to the model. Analysis of the survey data revealed that risk perception of driving situations is non-linear in the extremities of risk, showing that the input image are perceived as normally distributed instead of uniformly distributed.

The comparison further shows that a model with features and weights solely based on literature is only slightly capable of predicting the risk of situations with a Pearson correlation coefficient with the survey responses of 0.28, whereas a model with feature weights optimised is moderately capable of predicting the risk of situations with a correlation coefficient of 0.57. However, multivariate regression is better capable of predicting risk with a correlation coefficient of 0.70, and shows that features and weights based on literature were not enough to establish an accurate model. The features that have the most impact on the result are the information about other road users' location and heading, the ego vehicle velocity, and the road type.

Further research should focus on determining if speed is the cause of the perceived risk, or if it follows from other unknown predictors. Furthermore, an extension of the questions asked of the participants and the usage of videos instead of images can clarify the discrepancy between the literature-based model and the perceived risk of the participants.

Contents

1 Introduction

Automated vehicles are emerging. How fast the number of automated vehicles grows is still unclear, but according to Milakis et al., fully automated vehicles will become commercially available in the Netherlands between 2025-2045 [1]. It is still unknown if all vehicles will become fully automated, but if they do, there will be a change in the way the road users interact with each other. This change is most prevalent when human drivers and automated vehicles are mixing in traffic. For safety purposes, it is important that human drivers understand how automated vehicles work, but it is also important that automated vehicles understand how humans behave in traffic in order to predict their behaviour.

One part of human behaviour in traffic consists of rules that are applied in geographic areas, the formal rules. 'Stop before a red light' and 'Traffic coming from the right has the right of way' are two of the many rules in the Netherlands (and many other countries). The other side of human behaviour is not explained by general written rules, but by behaviour that can differ from human to human, the informal rules. Examples are a wave of the hand to thank for receiving the right of way, or yielding when you have right of way [2], both subject to regional differences [3] [4]. A part of getting a drivers licence in most countries is the answering of threat perception questions. These questions contain hints of events happening around the vehicle, and ask the participant what action to take. A ball on the street, for example, is an object that a human driver and an automated vehicle will avoid, but the human will also interpret the ball as an extra risk, because of the possibility of an emerging child. This is an informal rule, as there is no legal consequence.

This threat perception differs from hazard perception, and we can distinguish between them in the sense that threat perception includes dangers not currently occurring, whereas hazard perception includes the dangers currently occurring. Borowsky et al. mention hazard perception of drivers, where they study the differences in perception between age and skill groups [5]. The difference in perception between age and skill is also shown in similar studies that focus on hazard perception in the context of driver examination [6] [7]. Williams et al. examined drivers that already obtained a license, where they ask about the precautionary actions that drivers take to reduce risk. The found that defensive driving and obeying traffic laws were the actions that the highest percent of respondents took to reduce risk [8]. More sensor based approaches are performed by Okamoto et al., who estimated threat based on drivers intentions, while Kolekar et al. estimated threat based on objects [9] [10].

While these studies show that a significant correlation can be found between hazard/threat perception and age/skill/gender or could model risk by inserting threats in driving scenarios, all researchers determine the hazardous scenarios by themselves. They provide situations with chosen hazards, and test for differences in response or response time. Because the hazard that is perceived is already thought of by the researchers conducting the study, it is unknown why the researchers, or more generally human drivers, perceive that situation as hazardous.

Nuñez Valesco et al. approached this problem from the perspective of Vulnerable Road Users (VRU's) [11]. They found that the most influential factors that determine the perceived risk, or probability of adverse effects when exposed to harm were the speed of the vehicle, the gap size between the pedestrian and the vehicle, and the presence of a zebra crossing. But what are the influential factors from the perspective of a driver? Some factors are found by Malta et al., where they correlated the perceived risk to braking and verbal action. They found that

dangerous locations, busy intersections for example, could provide an estimate of the perceived risk [12]. Nuñez Valesco et al. used a theoretical framework explaining the way automated vehicles could affect VRU's to devise a list of potentially important variables. We will use a similar framework to devise a list of variables that could affect the driver risk perception. The framework that we will use is the parameterisation of threats, created by Breznitz in 1984 [13]. This framework is not based on environmental variables that influence the driver directly, but on a general concept of human threat perception. To use this framework, we have to connect the machine based sensory information to the human threat perception variables.

2 Method

Drivers' risk perception can be examined in multiple ways. We will focus on the creation of a qualitative model of human risk perception that is based on a human psychological model, and a quantitative survey where multiple participants are asked to determine risk in a certain traffic situation. The model predictions and survey results are compared to determine the most important features that influence the threat perception from the perspective of the driver. But before we can design the survey and model, we need a dataset that can provide relevant features.

Dataset

We opted for an existing dataset because the scope of this research is not the dataset, but the risk perception of participants and the corresponding features. The existing datasets are reviewed on the availability of information on the vehicle that collects data (the ego vehicle), VRU's, other road users, and road types. First of all, a dataset must contain VRU's and should contain data in the visible light spectrum. Six datasets contain VRU's in the visible light spectrum [14]: Caltech [15], KITTI Vision Benchmark [16], CityScapes [17], Tsinghua-Daimler Cyclist Benchmark [18], ETH [19], TUD-Brussels [20], PASCAL-VOC [21]. Only KITTI and Tsinghua-Daimler contain both pedestrians and cyclists. Although the Tsinghua-Daimler set contains VRU's, it does not contain cars, trucks, etc. The KITTI dataset is therefore chosen.

Type of input data

The KITTI dataset contains multiple videos of drives taken by their vehicle, equipped with cameras/laser scanners and GPS [16]. Moran et al. found that risk perception surveys are usually performed with videos, images, a simulator, or real-world test drives [22]. A dataset is used, therefore excluding simulator and real-world tests. Vlakveld examined both video and image test types for the Dutch drivers' license examination [6]. He found that both tests can differentiate drivers on accident risk, where young novice drivers with reported accidents performed significantly worse than young novice drivers without reported accidents. There was a difference between both tests when comparing the age of drivers, where the image test provided a clear increase in score with an increase in age. This could be explained by the fact that the video test requires more explanation. Vlakveld mentioned that for older drivers with little knowledge of computer games it was not immediately clear how to proceed with the experiment. The image test type is therefore chosen because we use an online survey that provides limited opportunity for extended explanation.

Tracklet parsing

In the KITTI Vision Benchmark, there are two methods of representing the object data: Per image object information (.txt files) and per object information (.xml files). The per object method uses one file per video and is based on the notion that the same objects can be in multiple frames. The per image method lacks this extra information, and is, therefore, unable to provide object continuation information over multiple frames. It is, however, the method that we

will use since single images are taken as the model input. The raw KITTI dataset contains per object information. This means that the per object information format should be parsed to the per image format. Christian Herdtweck (MPI Tuebingen) provided code to obtain Python data structures from the per object files on the KITTI dataset website [23]. These data structures are then parsed to the per image format.

Sample selection

After the object data is correctly formatted, we select a representative set of images to use for testing and validation of the model. A selection of the images is needed because there is a maximum number of participants due to monetary constraints (The Appen platform provides compensation of 25 cents for participants that finish the survey, this results in around 2000 participants for this study). These participants can only rate a certain number of images. This means that there that is a maximum number of images that can be rated, to ensure multiple raters per image. The right balance between raters per image and maximum images was found when the number of images was 210. We examine two selection methods, a random image selection and a uniform image selection.

Random image selection is initially the preferred method, because it removes selection bias. To randomly select the images, there are three selection steps. A random selection of a road type is first. The road type is based on the KITTI dataset and is 'city', 'residential', or 'road'. After the road type selection, there is a random selection to determine which drive is used. The last random selection is the selection of the frame of the video. With this method, an equal number of samples per road type is approximated. However, the number of images sampled per video can differ, because of the different number of drives per road type in the KITTI dataset. To ensure repeatability, the randomisation seed is saved.

In contrast with random image selection, uniform image selection does not utilise any form of randomisation. Every road type produces the same number of samples. The sample index is determined by equal frame spacing. The spacing itself is determined by dividing the total number of frames of a road type by the required number of samples.

The frames are all part of videos. This raises a problem for the random image selection method. Multiple frames are close to each other in the video time-wise, and some frames were therefore too similar, shown by the example sampling of Figure 1. The uniform image selection method maximises the temporal distance between the frames, resulting in less similar frames. The creation of diverse samples is the reason uniform image selection is chosen as the selection method.

Dataset features

The dataset contains information about the objects surrounding the vehicle and the vehicle itself, but it is possible that these are not all applicable as features for a model. First, we examine the raw information, and then select and convert relevant features from that information. The first feature that we find does not need conversion: the velocity of the vehicle.

The next features are based on the objects that the vehicle detects. The objects contain the

size, location, occlusion (partly visible), truncation (partly out of image), and observation angle (alpha) information. The observation angle alpha is: $r_y - \theta$, where θ is the view angle of the object in the camera image, and r_y is the angle of the object frame compared to the camera frame. These features can be subdivided into the mean, minimum, maximum, and sum of all objects in an image. Furthermore, a general count of objects is added as a feature.

The third type of feature is based on the road type of the image. This is provided in the dataset, as the raw KITIIT Vision Benchmark data is divided online into six categories. Three categories are chosen: City, Residential, and Road. Because three categories are chosen, the total number of images should be divisible by three to ensure that each category contains an equal number of images.

Finally, there is information in the images that are not available in the dataset but can be manually annotated by the author. These are the number of braking cars (brake lights), cars moving towards the vehicle (front visible), and cars moving away (back visible). This results in a list of 22 features, shown in Table1. The table also shows the mean and standard deviation of all features, to provide an overview of the order of magnitude of the features.

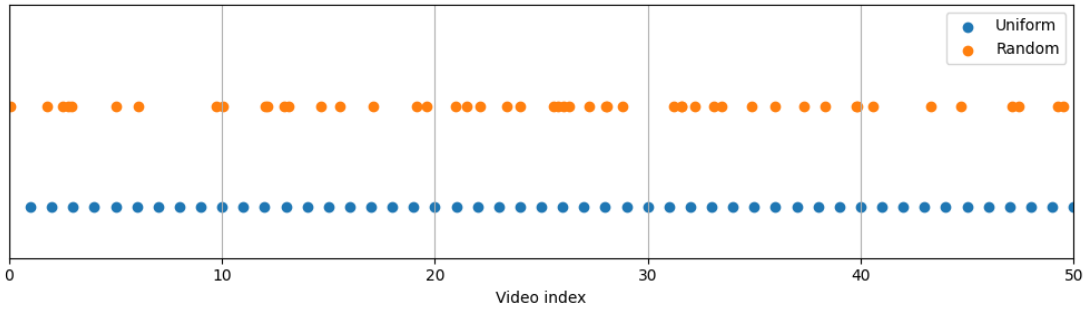

Figure 1: Difference between random and uniform sampling

Table 1: features available from dataset, where object means another road user in the image.

feature	Description of feature	μ	σ
general_velocity	Velocity of the ego vehicle [m/s]	9.05	5.37
general_distance_sum	Distance to all objects summed	21.23	18.64
general_distance_min	Minimum of distance to all objects	3.51	2.46
general_distance_mean	Mean of distance to all objects	4.45	2.71
general_number_objects	Number of objects in image	4.40	3.92
manual_car_toward	Number of cars with headlights visible	0.64	1.02
manual_car_away	Number of cars with taillight visible	0.89	1.62
manual_brakelight	Number of cars with brake light visible	0.22	1.00
alpha_min	Minimum observation angle of objects [-pi/pi]	0.09	0.13
alpha_mean	Mean observation angle of objects [-pi/pi]	0.24	0.18
alpha_max	Maximum observation angle of objects [-pi/pi]	0.46	0.39
occluded_mean	Mean of occlusion state of objects	0.54	0.53
occluded_sum	Sum of occlusion state of objects	3.68	4.84
truncated_mean	Mean of truncation state of objects	0.13	0.21
truncated_sum	Sum of truncation state of objects	0.78	1.40
size_mean	Mean of all object sizes	17.56	25.07
size_max	Maximum of all object sizes	32.79	46.74
size_min	Minimum of all object sizes	10.84	22.77
size_sum	Sum of all object sizes	66.11	67.55
road_road	Boolean of road type road	0.33	0.47
road_residential	Boolean of road type residential	0.33	0.47
road_city	Boolean of road type city	0.33	0.47

3 Model

To be able to create a model that perceives risk, we need to extract objective features of this subjective concept. We can do this by basing risk perception on the occurrence of threats in the environment, where a threat is an indication of something happening. We use the parameterisation of threats by Breznitz to accomplish this [13]. The seven features devised by Breznitz are used as an input for a Multi-Criteria Decision Making Process (MCDMP). These features of Breznitz, and their derived model features and weights, will be examined next.

3.1 Parameterisation of threats

The probability of the impending threat in general

The general probability is based on the notion that there is a general probability to have a deadly accident when driving on the road. The differentiation of a deadly accident is used because that data is well documented while data on minor injuries is not well recorded. This general probability is often expressed as a micromort or an accident occurrence per driven kilometre. The data provided for the Netherlands by the Institute for Road Safety Research (SWOV) is different and shows accident occurrence per kilometre of road length. This data also shows that there is a different accident occurrence for different road types [24]. That is why different weights will be used used for different road types. To obtain a probability for a certain road, we divide the casualties by the total road length of that road type [25]. We obtain the chance of deadly accident per kilometre road length in 2018 as: municipal = $\frac{404}{126458} = 0.32\%$, provincial = $\frac{105}{7745} = 1.36\%$, and state = $\frac{81}{5384} = 1.50\%$. The accidents occurrence per kilometre does not take into account the traffic density of a certain road. This could influence the chosen weights, but we could not find corresponding traffic density data.

These road types can be coupled to the road type markers used in the KITTI dataset, where the road types *city* and *residential*of the KITII dataset correspond to the municipal data, and the road type *road* corresponds to the provincial and state data.

The type of threat

When driving a car, the driver encounters multiple different types of road users and objects. Bazilinskyy et al. showed that there is a difference in perceived risk, especially when encountering multiple pedestrians and cyclists in a busy city [4]. It is therefore expected that there will be different risk perceptions for different road users, and furthermore, we expect pedestrians and cyclists to have a high risk-perception factor. The data of the SWOV shows a similar result, where casualties of traffic accidents are mostly pedestrians and cyclists [24]. This study shows, however, that there is one other road user with a high amount of casualties: another car. The risk perception of other cars may be low in the other studies because drivers expect more cars on the road than VRU's. This makes the appearance of VRU's on the road unexpected. We have therefore chosen for a low perceived risk for cars, and a high perceived risk for VRU's. Road users like trams and trucks receive an increased amount of perceived risk when their occurrence on the road decreases. We will take 1 as a maximum weight, and 0 as a minimum weight with

steps of 0.2. The road users that are available from the annotated KITTI dataset are ordered by expected appearance on the road [25], and supplied with the weights shown in Table 2. In this table, 'ped' means pedestrian, 'pedsit' means sitting pedestrian, 'cycl' means cyclist, 'misc' means miscellaneous, and 'dc' means don't care, which is a label given to objects that are far away.

Table 2: Model weights for different road user type.

car	van	truck	ped	pedsit	cycl	tram	misc	dc
0.2	0.4	0.6	1	0.2	1	0.6	0.2	0

The magnitude of the potential harm

We define the magnitude of potential harm as the magnitude of potential physical damage. The magnitude of the potential harm is based on the amount of kinetic energy in a collision [26]. Thus, heavy vehicles have higher potential harm to the driver than light vehicles, and fast-moving vehicles have higher potential harm than slow-moving vehicles. The weights chosen for road users corresponds to this notion except for VRU's. The magnitude of potential harm lays here with VRU's and not the vehicle, which results in large potential harm, not for the driver, but the VRU's. These risk weights correspond to the weights established in Table 2, and they are therefore kept the same.

The imminence of the threat

The imminence of a threat is a factor of object location and movement compared to the ego vehicle location and movement [27]. The most used measure of imminence is time to collision. To calculate the time to collision, the velocity and direction of all road users are needed. The KITTI dataset does not contain this information directly, only indirectly by temporal analysis of point cloud data. This analysis is not within the scope of this research. Time headway is therefore chosen instead of time to collision. This has the consequence that it cannot distinguish between imminent danger and the potential danger [28].

The extent to which the threat can be dealt with

The extend to deal with threats can be expressed as the possibility of the driver and his vehicle avoiding a collision. This is a combination of vehicle dynamics and the driver's control response. Both are unknown and require a model to estimate the possibility to avoid. Because pictures are used for the risk perception experiment, both models are not implemented. However, it is assumed that the participants can obtain some information about collision avoidance from the images. Jurecki and Stańczyk found that the percentage of drivers that respond with actions, compared to an increased time to collision scales nonlinearly [29]. This feature will therefore extend the imminence feature (time headway) with a quadratic function as shown in Equation 1.

$$imminence = \sum_{n=1}^{n} w_m (\frac{d_n}{v})^{\frac{1}{b}} \tag{1}$$

Where n is the object in the image, d_n is the distance to that object from the camera, and v is the velocity of the vehicle. The weight is initially chosen as $w_m = 1$ and the shape factor as $b = 0.5$, because no further information is available.

The extent to which similar threats have been experienced before

The experience of threats is a user-specific feature. Although it is possible to estimate the possibility that a threat is experienced before by the general probability of the threat, the general probability is already used as a feature by the general probability of a threat.

The level of confidence of the observation

The annotated data from the KITII-dataset has a ground truth. It can be useful to implement a confidence feature when the model is used in a real-life scenario, with confidence levels of object detection methods. With the ground truth of the KITTI dataset, the confidence is always 100 per cent and therefore omitted.

3.2 Combining features

The features as described above result in three main groups of features: road user type t, imminence m, and road type r. For the objects in an image n, the corresponding object weight w_t is summed up. The road type r is simply the weight of the road type. The imminence is taken from Equation 1. These individual features are summed up in the end, shown in Equation 2.

$$R = \sum_{n=1}^{n} (w_{t,n} + w_m (\frac{d_n}{v})^{\frac{1}{b}}) + w_r \tag{2}$$

Where R is the predicted risk per image, $w_{t,n}$ the weight of a road user type of an image, w_m is the imminence weight, and w_r is the weight of the road type of an image. R will also be referred to as model_combination.

3.3 Optimisation method

The feature weights are optimised after the weight factors of the features determined with the parameterisation of threats are examined. There are 14 features to be optimised for the model. This requires an optimisation method capable of optimising multiple features. Furthermore, the optimisation method should be able to not only find the local minima, but also the global minimum. The python library *SciPy* contains 5 global optimisation methods [30]. Because of

the large feature space and the inability to calculate a gradient, the differential evolution method is chosen. The population size is relatively small, so the Exp1bin strategy is preferred [31].

4 Survey method

To validate and optimise the model, an online survey was constructed and distributed via the Appen platform [32]. In this survey, the participants first answer general questions about age, current lighting conditions, etc (see full survey in appendix E). After the general questions, the drivers are asked similar questions to the 'Driver Behaviour Questionnaire' as used by De Winter and Dodou [33]. The general questions are included to allow for filtering on specific respondent traits if necessary. At the end of the questionnaire, the participants are asked to follow a link to a web page hosted on the Heroku Cloud Application platform [34]. On this web page, the participants are asked to move a slider to their perceived risk value on a continuous scale from 0-10 for a subset of 100 out of the 210 selected images of the KITTI dataset. An example is shown in Figure 2. Although the participants rate on a scale from 0-10, the data collected is the response of the participant times 10, resulting in responses from 0-100. After the grading of the images, the participants are given a unique code that they fill in on the Appen platform survey.

Figure 2: Example of survey question

The subset of 100 images selected from the KITTI dataset is determined by the average time that a participant needs to finish the survey. The aim for a time that a participant on average needs is 30 min, to ensure proper engagement of the participant. From previous studies, the number of images per participant should be 100 to result in an average time of 30 min.

Removal of cheaters

The Appen platform provides compensation for participants that finish the survey. This compensation can be a source of income for participants, especially where the compensation approaches the hourly wage. This creates an incentive for participants to fill in the survey multiple times, without the need to grade the images. To counter this, we filtered the responses of participants and excluded responses where the unique token was used multiple times, keeping only the

first response. The participants that used this method were reported to the platform, and new participants were selected.

Restructuring survey results

The responses of the participants on the Heroku part of the experiment are saved after each image. These responses are saved with the meta-data of the participant to a JSON file. All individual responses are then put together by the respondents' unique code, which should result in 100 responses per participant.

Matching unique codes

The participants received their unique code at the end end of the Heroku part of the experiment and submitted their unique code at the Appen part of the experiment. This makes it possible to combine both parts of the survey. After filtering and combining, the survey yielded 1918 unique responses.

5 Survey analysis

Of the 1918 participants, 1447 rated all 100 images, 145 rated 99 images, and 77 rated 98 images. The number of images rated per 10 image intervals is shown in Table 3. Data acquisition overflow, where multiple participants send data at the same time to the acquisition server is probably the cause of some dropped responses. Dropped responses will not provide a significant difference in the result because of the high amount of rated images per participant.

Table 3: Amount of responses per participants per 10 image interval

# Images	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
Participants	1	0	0	1	0	3	9	15	48	1841

The average age of the respondents was 35.5 years old. The oldest participant was 85 years of age, and the youngest participants were 18 years old. The mean of all responses $\mu = 33.28$, with an average standard deviation of all responses of $\sigma = 25.89$. The mean and standard deviation of the responses for individual images sorted by their mean response are shown in Figure 3, where the perceived risk is the response of the participants. The figure shows a linear increase of mean perceived risk responses in the middle and an non-linear decrease/increase at the sides. This can indicate that perceived risk of the images used is normally distributed instead of the initially assumed uniform distribution. To test this, we will predict a normal distribution based on the mean and standard deviation of the results. The results and predicted results are shown in Figure 4. The figure clearly shows that the images are not perceived as a uniformly distributed dataset, but instead like a normal distributed dataset. With the responses of the participants gathered, the next step is a general analysis of the data to determine the reliability. We use three methods of checking the reliability: autocorrelation, manual observation, and grouped responses.

5.1 Autocorrelation

First, we will check the reliability by performing a correlation between the first half of the participants, and the second half of the participants. If these two halves replicate, it is an indication that the data is reliable. The two halves are sorted by the unique code that the participants received during the experiment. Both groups have 859 participants with group 1 a mean of $\mu_1 = 32.56$ and a standard deviation of $\sigma_1 = 26.99$. Group 2 has a similar mean of $\mu_2 = 33.84$, with a standard deviation of $\sigma_2 = 27.03$. The correlation of the first half with the second half is $r = 0.98$. This is a strong correlation between the first half and the second half of the participants, shown in Figure 5. The data also shows large standard deviations in the responses. This is to be expected because the question: 'how high is the risk?' is subjective. One person might find risk zero when there is no obvious hazard, while another person might find the risk 40 per cent for the same situation.

The deviation between participants is also shown when analysing the correlations of the participants themselves with all other participants. These correlations have a mean of $\mu = 0.38$, a median of $\bar{x} = 0.43$, a standard deviation of $\sigma = 0.22$, and a maximum of $M = 0.80$. There were some participants that did not correlate with the average responses at all ($r \approx 0$). This can indicate that these participants have either a completely different risk perception than the other participants or, more likely, they were not genuine with their responses. Even with these large standard deviations, the averages replicate.

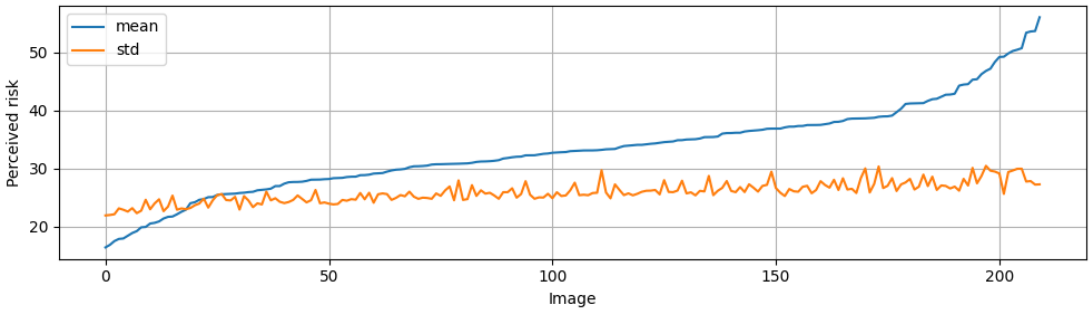

Figure 3: Mean and standard deviation of survey responses, sorted by perceived risk.

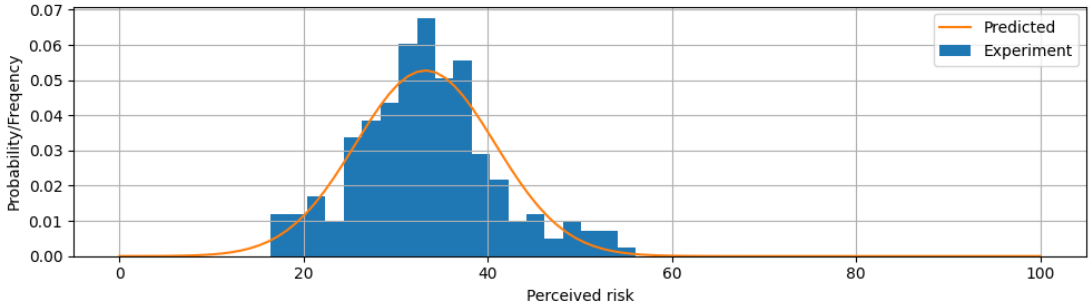

Figure 4: Estimated and resulting distribution of perceived risk.

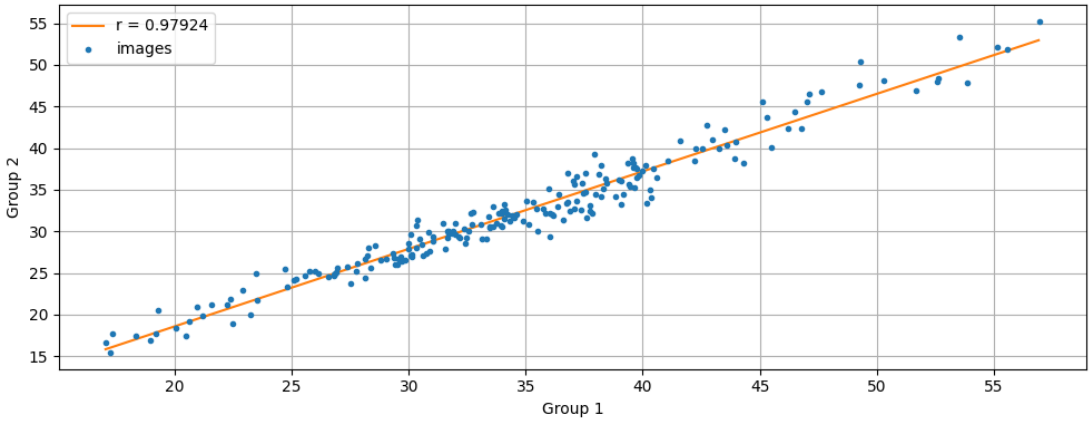

Figure 5: Scatter of responses per image of group 1 and group 2.

5.2 Manual observation

The second method of checking the reliability is a manual check of the general trend by ranking the images by risk. The top 10 least risky images are all images on a similar road with few other road users and clear vision. The top 10 riskiest images are from situations where there is uncertainty in the image. The images that score highest on risk contain multiple road users and it is unclear what the other road users are doing. These results are to be expected, for more road users in a situation are expected to provide a higher risk. The most and least risky images, as ranked by the respondents, are shown in Figure 6 as an illustration.

Figure 6: Images ranked by respondents, with least risk (top), and most risk (bottom).

5.3 Grouped by age/driving experience

The final analysis is a grouping of participants by four risk predictors of the 'Driver Behaviour Questionnaire' as used by De Winter and Dodou [33]. The four predictors used are: disregarding safe distance (*close*), disregarding hands-free phone usage (*phone*) disregarding speed limit (*speed*), and accident occurrence (*accident*). The questions asked for the predictors are shown in Table 4. Correlating these predictors with the survey risk score provides a negligible correlation coefficient ($r < 0.025$). This means that these features are not an adequate predictor for risk perception. The inter-group correlation on the other hand shows similar results as De Winter and Dodou found in the literature, with correlations between 0.33 and 0.49, as shown in Table 5 [33]. This supports the notion that these predictors of the driver behaviour questionnaire can significantly predict accident occurrence, but not risk perception. Furthermore, the inter group correlations show that different types of risk taking behaviour are significantly correlated.

Table 4: The four questions asked to obtain the predictor variables.

Predictor variable	Full question	Choices
speed_occurence	How often do you do the following?: Disregarding the speed limit on a residential road	0, 1-3, 4-6, 7-9, or 10 times per month.
close_occurence	How often do you do the following?: Driving so close to the car in front that it would be difficult to stop in an emergency	0, 1-3, 4-6, 7-9, or 10 times per month.
phone_occurence	How often do you do the following?: Using a mobile phone without a hands-free kit.	0, 1-3, 4-6, 7-9, or 10 times per month.
accident_occurence	How many accidents were you involved in when driving a car in the last 3 years?	1, 2, 3, 4, 5, and more than 5 times.

Table 5: Inter group correlations.

	(1)	(2)	(3)	(4)
speed_occurrence (1)	1.0			
phone_occurrence (2)	0.42	1.0		
accident_occurrence (3)	0.33	0.33	1.0	
close_occurrence (4)	0.49	0.44	0.38	1.0

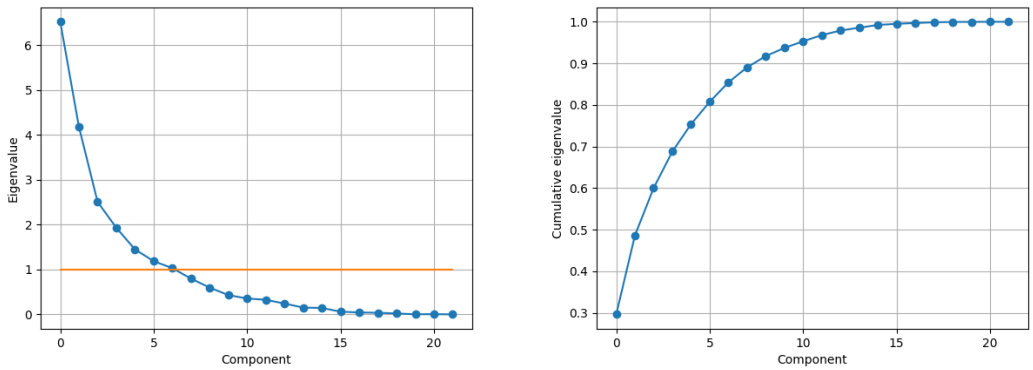

Figure 7: Scree plot and cumulative eigenvalues of eigenvalues.

6 Benchmark

With the reliability of the survey data established, the next step is setting a benchmark for the model. The method used to provide a benchmark is multivariate regression. This is performed on the features available from the dataset. Afterwards, a principle component analysis will be performed to check for significant pixel-based features. This can indicate if relevant image features are missing.

6.1 Multivariate regression

Multivariate regression can be used to predict the outcome of the experiment. Normally, the principal components are used for a multivariate regression. We chose to use predictor variables so we can also obtain the impact that predictor variables have on the prediction in the form of standardised linear equation parameters (β_{stand}). The relative size of the parameters resulting from a multivariate regression indicate their influence on the result, and will in our case indicate what variables make a situation risky. To perform multivariate regression, it is necessary to determine the relevant predictors. Twenty-two features can be used, but not every feature is a good predictor. To determine which features are good predictors, we calculate the eigenvalues and the corresponding eigenvectors of the features (Figure 15 in the appendix). The eigenvalues are sorted based on their values and shown in Figure 7 as a Scree plot. This will ensure that the variables with the most variation are used in the multivariate regression.

The Scree plot shows that the influence of the eigenvectors diminishes fast. The "elbow" of the plot is between eigenvalue 5 and 8, which can be an estimate of the number of features to use. If we use the Kaiser Rule, there are 7 eigenvalues larger than 1, which is similar to the estimate from the Scree plot. To obtain the predictors from the features, we can use the cumulative eigenvalues. The first 6 eigenvectors contain at least 80 per cent of the variation of the data.

The features which have values higher than 0.4 in these 6 eigenvectors are taken as predictors. This results in the 8 predictors shown in Table 6. The threshold value of 0.4 is chosen to result in 8 predictor features. Both the Kaiser rule and the Scree plot method can overestimate the number of features needed, although this is less likely with the large sample size used here [35]. Furthermore, using predictors from the dataset instead of the principal components will always result in the necessity of an equal amount or more predictors to obtain the same variance.

With these predictors, linear multivariate regression is performed. The first step is a standardisation of the predictors. After the standardisation, the linear multivariate model is fitted to the first half of the responses (image index 0-104). This fit is done using the SciPy sklearn linear regression fitting which uses Ordinary Least Squares [30]. The values shown in Table 6 are the values of the predictors without the bias/interception added, resulting in the relative influence values. This results in a correlation of $r = 0.70$. This is a strong positive correlation, with the general_velocity predictor standing out because of the higher absolute β_{stand} This regression will be the benchmark for the model.

6.2 Principal component analysis of the image

The linear multivariate regression model that was previously fitted on the data, is fitted on object features. There are, however, more features which could possibly be important. These are the raw image features or, in other words, the pixels. To check if the pixel features are important, a principal component analysis is performed on the image pixels. The images of the dataset are first scaled down to 621x188 pixels from their original amount of 1242x375 pixels. The amount of components chosen for the analysis is 50, which is chosen as a balance between the number of components and computing time. All colour channels are analysed separately, plus an additional grey channel. The resulting principal components are correlated with the survey data, resulting in the correlations shown in Figure 8.

The figure shows that there are several components with a small correlation. For some components, for example, component three, all colour channels have roughly the same correlation. This indicates that the principal component is an illumination component. Other components, for example, 16, show a difference in positive and negative correlation per colour channel. This indicates that the principal component is a colour component. Principal component three is shown as an eigenimage in Figure 9. The white section at the bottom of the image indicates that the component is responsible for the variance of the road illumination. A possible reason for this correlation could be that when vehicles are close, the illumination of the road in front of the vehicle changes because a vehicle obstructs the view of the road. It is important to note that the principal components with high correlations could be a valuable addition to the multivariate regression. The downside to using these features is that every eigenimage needs interpretation of what the image represents. This is increasingly harder to do with increasing eigenvalues. Eigenimage with eigenvalue seven is similar to eigenimage with eigenvalue three, but it is already difficult to interpret the shapes in the image.

Table 6: Predictors of multivariate regression, with their corresponding coefficient, mean of predictions, standard deviation of predictions, confidence interval of predictions ($\alpha = 0.05$), and scale. The training set was the first half of the images, sorted by index (0-104). The testing set was the remainder (105-209).

	Predictor	β_{stand}	μ	σ	Upper	Lower	Scale
Training	manual_brakelight	1.6351	0.4200	1.8404	0.7737	0.0663	Ratio
	size_mean	-1.0272	-0.7543	1.1293	-0.5373	-0.9713	Ratio
	alpha_min	0.9210	0.6421	0.8944	0.8140	0.4702	Ratio
	size_min	0.5618	0.3023	0.6212	0.4217	0.1829	Ratio
	road_city	1.1061	0.8492	1.1330	1.0669	0.6314	Binary
	general_velocity	-3.7882	-6.1793	3.7090	-5.4665	-6.8921	Ratio
	manual_car_toward	1.1105	0.7781	1.2431	1.0170	0.5392	Ratio
	road_residential	0.7841	0.5861	0.7984	0.7395	0.4327	Binary
Testing	manual_brakelight	1.6351	0.3111	1.4162	0.5833	0.0389	Ratio
	size_mean	-1.0272	-0.6879	0.9237	-0.5104	-0.8655	Ratio
	alpha_min	0.9210	0.6041	0.9550	0.7876	0.4205	Ratio
	size_min	0.5618	0.2338	0.4992	0.3297	0.1378	Ratio
	road_city	1.1061	0.7151	1.0852	0.9237	0.5065	Binary
	general_velocity	-3.7882	-6.6278	3.8884	-5.8805	-7.3752	Ratio
	manual_car_toward	1.1105	0.6225	0.9656	0.8081	0.4369	Ratio
	road_residential	0.7841	0.5227	0.7758	0.6719	0.3736	Binary

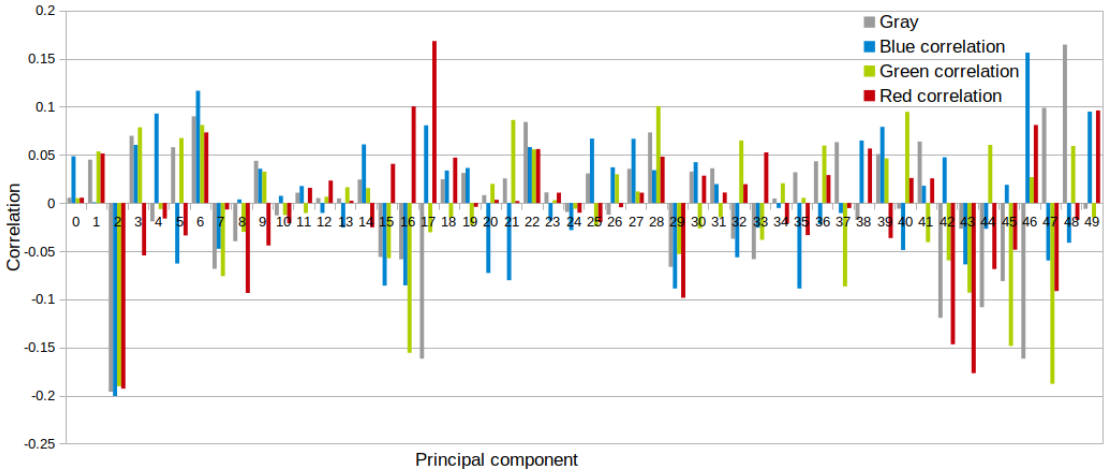

Figure 8: Correlations for the eigenvalue for each eigenimage.

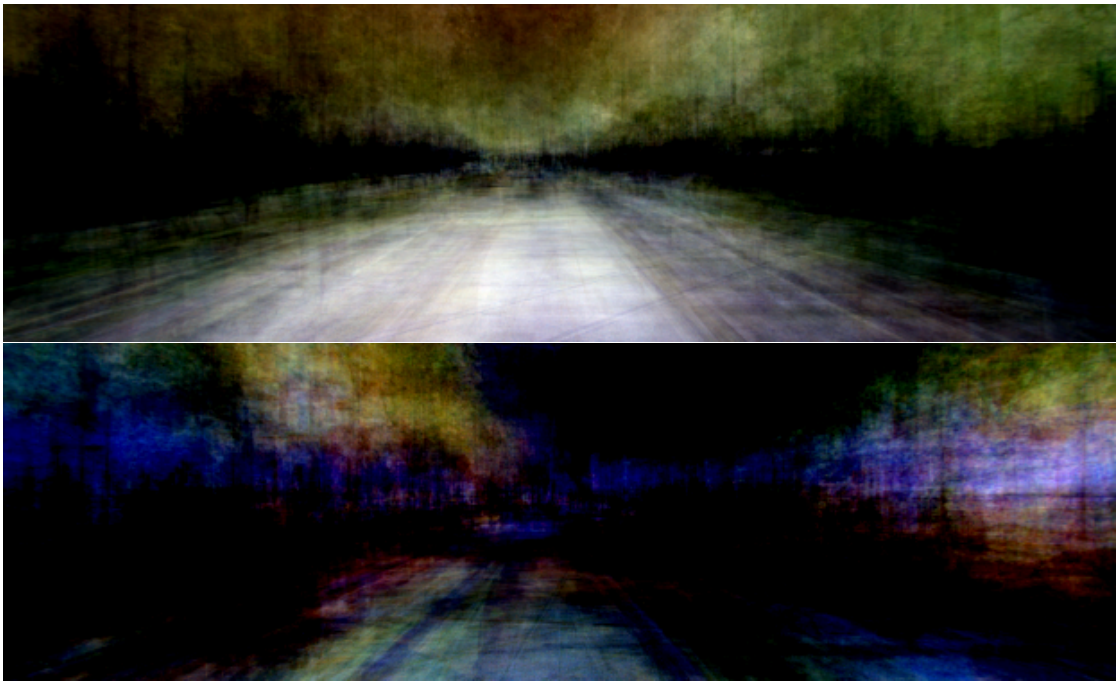

Figure 9: Eigenimage with third principal component (top), and seventh principal component (bottom).

7 Model Results

After the creation of the benchmark, we analyse the results of the literature-based model. The correlations of the model with the participants responses are examined first, then we will examine the images ranked by the model.

7.1 Correlations

The output of the model with non-optimised feature estimates as described by Equation 2 is correlated with the responses of the participants and all individual features. The resulting correlation is $r = 0.28$ for the model combination (R), $r = 0.47$ for the model type (t), $r = 0.28$ for the model imminence (m), and $r = -0.34$ for the model probability (r). These are small to medium correlations, and not as high as the benchmark ($r = 0.7$). On overview of all correlations is shown in appendix B.

The literature-based model is now optimised, with bounds at 0 and 10 for every feature this results in $r = 0.54$ for the model combination, $r = 0.48$ for the model type, $r = 0.35$ for the model imminence, and $r = 0.34$ for the model probability. These are medium correlations, with the combination correlation being the highest, and the probability correlation now being positive. The model is optimised again, but this time with bounds chosen at -100 and +100. This results in $r = 0.57$ for the model combination, $r = 0.5$ for the model type, $r = 0.27$ for the model imminence, and $r = 0.35$ for the model probability. This final optimisation shows an increase in the model_combination correlation, but a decrease in type correlation.

Table 7: Model predictor variables before optimisation (Init.), after optimisation (Opt.), and Pearson's correlation coefficient of the model with the survey results.

	feature	Init. Estimate	Opt. Small bounds	Opt. Large bounds
w_t	type/car	0.20	0.00	9.75
	type/van	0.40	6.90	87.3
	type/truck	0.60	4.71	45.1
	type/ped	1.00	6.37	94.1
	type/pedsit	0.20	8.69	99.6
	type/cycl	1.00	3.47	83.2
	type/tram	0.60	0.00	-100
	type/misc	0.20	5.59	94.6
	type/dc	0.00	7.35	-28.3
w_r	prob/city	0.32	9.76	44.1
	prob/residential	0.32	7.26	27.2
	prob/road	1.43	0.59	-85.8
w_m	imm/gain	1.00	0.86	0.00
	imm/bias	0.50	5.09	0.09
r	model_combination	0.28	0.54	0.57
	model_type	0.47	0.48	0.50
	model_imminence	0.28	0.35	0.27
	model_probability	-0.34	0.34	0.35

The feature values of the estimates, optimisation with small bounds, and optimisation with large bounds are shown in table 7, together with the correlations of the model, where r is the correlation. The result that is the most interesting in the initial estimate is the negative sign of the model_probability. This indicates that participants find city and residential roads more risky than other roads. This is contrary to the initial estimated, where the speed of the vehicles was assumed to influence the perception of risk more. This unexpected negative weight of the model_probability, which is not accounted for in the initial estimate, is accounted for when the model is optimised. The weight for the road types are adjusted, which result in a similar, but positive, result.

Even after optimisation, the model still performs worse than the benchmark. The chosen predictors in the benchmark differ substantially with the features used in the model, where manual_brakelight, size_mean, alpha_min, size_min, and manual_car_towards are absent. The only shared features are the road_type and general_velocity features. The difference between the model and benchmark is therefore probably due to the difference in feature selection method (literature based or principle component based), and not a difference in chosen weights or a different method of combining the features. The correlation between the optimised model with small bounds and the responses of the participants are shown in Figure 10. Although the optimised model does not outperform the benchmark, it does perform better than most random individual participants ($r_{participants} = 0.43 < r_{model} = 0.54$). This is shown in Figure 11, where the correlation of individual respondents with the other respondents is shown.

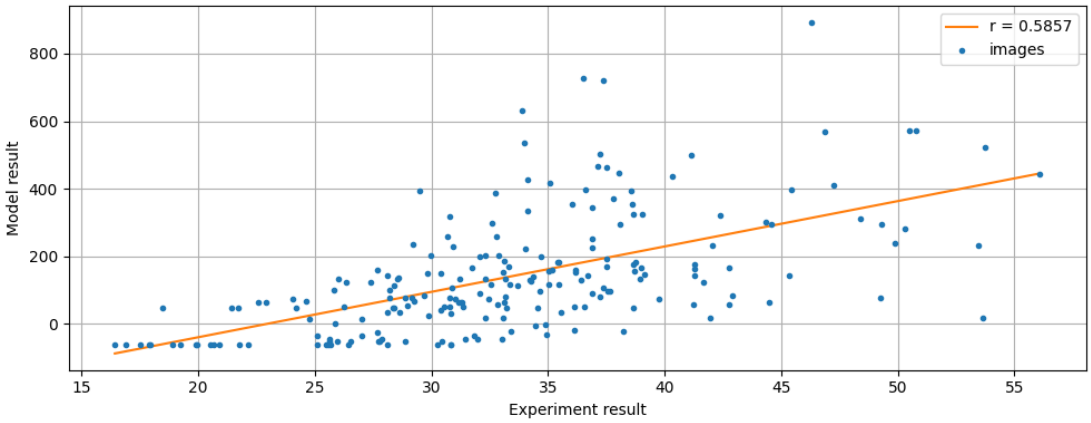

Figure 10: Correlation between the model and the experiment result.

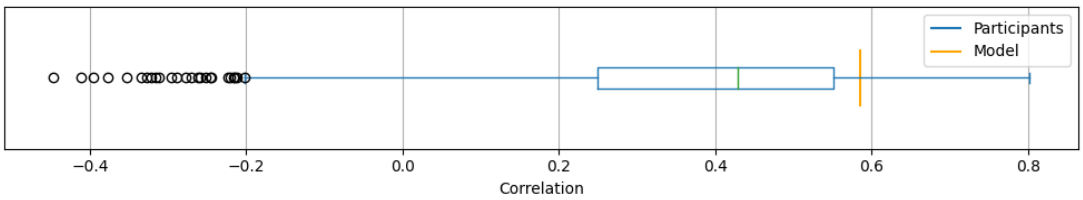

Figure 11: Correlations of individual respondents and model with all respondents.

7.2 Ranked images

The final optimised model did not exactly rank the images the same as how the participants ranked them, but there are image features that can be recognised in both the ranking of the model, and the ranking of the participants. The least risky images, as predicted by the model, are similar to the least risky images of the participants (figure 12 top). It is the same road with no other road users and a clear view as the participants ranked (figure 6 top). Furthermore, the most and least risky images as ranked by the participants, are found in the top and bottom 10% images ranked by the model respectively.

The image that was ranked third most risky by the participants, figure 13 top, is an unusual result in the model results: this image is the 161^{th} most risky image by the model. The difference in rank could be explained by the unpredictable movement of the garbage truck. It is also hard to see in the image if the vehicle is turning left already, or waiting on the truck. This explanation is supported by the images where the garbage truck is moving away, figure 13 bottom, where the further the garbage truck is around the corner, the lower the perceived risk. Showing that it is not the garbage truck itself, but the position of the truck compared to the vehicle.

Figure 12: Images ranked by model, with least risk (top), and most risk (bottom)

Figure 13: 3^{rd} Most risky image by participants, 161^{th} most risky by model(top). 105^{th} Most risky image by participants, 175^{th} most risky by model (bottom)

8 Discussion

The first question that follows from the results is: Why is a higher speed perceived as lower risk? We think it is unlikely a causation. It is, however, possible that both the higher speed and the lower risk perceived have a common cause. An example of this cause could be the environment where the vehicle is driving. A clear road, with no other road users and safety features facilitating higher speeds, could provide a safer perception. Whereas driving slow in a hectic city street could provide an unsafe perception, with a corresponding high perceived risk. The medium correlation of the road_type feature with the velocity hints that this could be indeed the case (appendix B). Further investigation of the velocity risk correlation reveals that road type does influence the strength of the correlation. The correlation is lowest in 'city' road types, and highest in 'road' road types, this is illustrated in Figure 14.

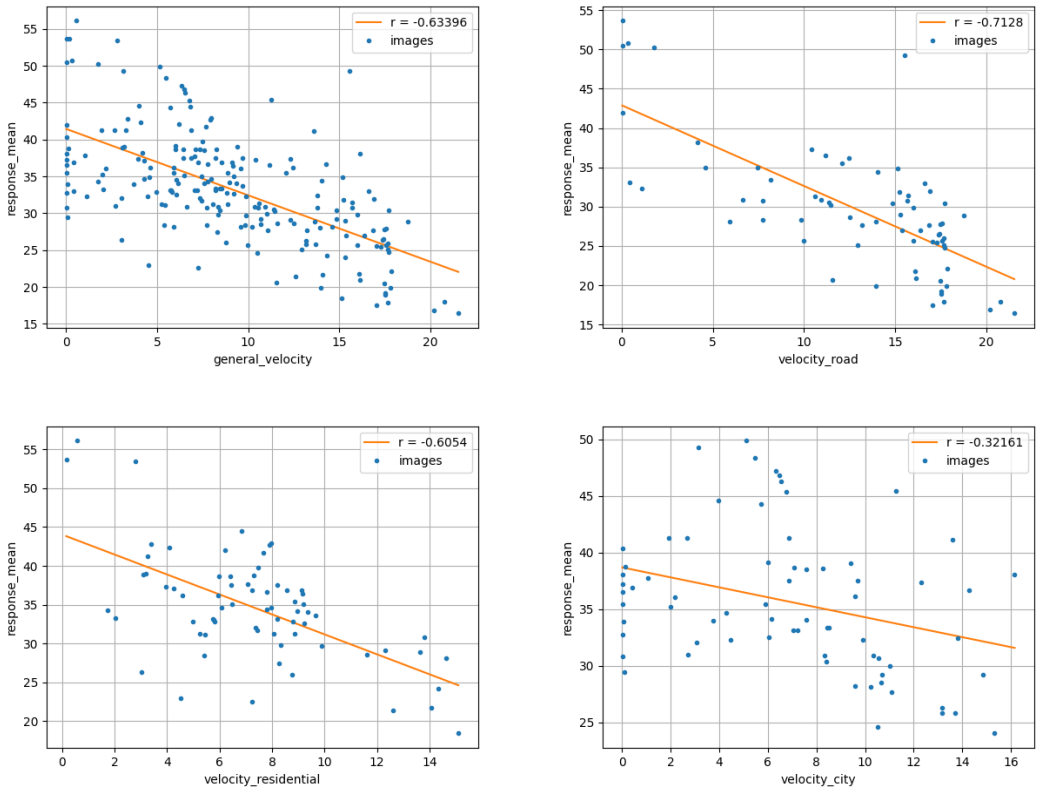

Figure 14: Correlation of velocity with perceived risk (response_mean), with separation by road type

Another possible explanation is the medium correlation between the number of road users and the velocity. Brackstone et al. found a negative correlation between the velocity of the vehicle and the amount of vehicles per kilometre [36]. We can check if that is also the case here, by analysing the correlation between the velocity and the number of vehicles. This correlation is $r = -0.32$ (appendix B), which is similar to the result found by Brackstone et al., but at the

same time a weaker correlation. This indicates that this cannot be the full explanation for this phenomenon.

It is possible that the results are influenced by the way the question about perceived risk was formulated. The question was: "As a driver, how risky would you judge this situation (0 = no risk, 10 = extreme risk)?". The term risky can be interpreted differently by different participants. We assumed that participants would interpret the question as: "How unsafe do you think this situation is?" or "How dangerous do you think this situation is?". This interpretation does not separate potential hazards and apparent hazards. Drivers could interpret the question to exclude potential hazards, and only include apparent hazards. This exclusion could explain why higher speed, and potentially less apparent hazards score lower, while low speed and more apparent hazards score higher.

The second question that follows is: Why are the features before the optimisation different from the features after optimisation (Table 7)? The model probability features show an inverse in magnitude, were the 'city' and 'residential' types increase and the 'road' type decreases. This is because the predictor variables for the model probability have changed. The change shows that the perceived risk in residential and city is higher than the perceived risk on other roads. This is conflicting with the deadly accidents per km road length from literature. This could be because of the usage of accident per road length, instead of per driven kilometre. The model type features also show a very different result after optimisation than the initial estimate. Cars, trams and 'don't care' seem to matter little when predicting perceived risk. For trams and 'don't care', this is most likely due to the fact that there are very little images ranked with a tram or 'don't care', resulting in overfitting.

9 Conclusion

The results of the online survey show that participants, in general, agree on risk perception. The survey also shows that the images provided for risk ranking seem normally distributed instead of uniformly distributed.

The model, with weights and features chosen by literature, is not capable of accurately predicting risk in images. When optimising the weights, this model is moderately capable of predicting risk in images, especially compared to individual participants. The best method however, is a multivariate regression, resulting in a correlation of $r = 0.7$. This shows that the model needs to be improved to achieve better performance. This could be done by using the features resulting from the multivariate regression.

The benchmark shows that features other than time to collision or time headway, can be used for risk perception models. This can be valuable in situations where information about other road users' movement and intention is not available. Furthermore, the predictors of the benchmark provide new information about the human perception of the surroundings while driving. Creating a better understanding why humans perceive risk the way they do, with information about other road users' location and heading, egovehicle velocity, and the road type having an important role.

10 Recommendations

To extend the contribution of this work, there are several recommendations. The first recommendation is a different style of survey. A ranking by comparison of images, where participants rank two or more images on risk at the same time can provide a support of this survey. Another recommendation is the usage of video instead of images. Both the image ranking and the use of video can provide further information on the large negative correlation of the velocity with the responses, by being able to separate apparent and non apparent hazards.

A second recommendation is an extension of the question asked of participants. An addition of a question "If you think this is a risky situation, explain why", could explain why certain images that seem similar are rated differently, for instance the garbage truck example. This can help determine why participants agree or disagree. Furthermore, extra questions that ask the same, but in different ways can provide more information an participants' consistency and may improve survey reliability.

A third recommendation is the usage of accident occurrence per driven kilometre instead of per kilometre road length. This will clarify the discrepancy between the literature-based model and the perceived risk of the participants.

11 Resources

The code written to perform the analysis, the resulting images, and anonymous participant data can be found online at: `https://github.com/jhoogmoed/HumanThreatPerception`. The code is split in three sections: KITTI dataset parsing, survey data parsing, and model comparison/optimisation.

References

[1] D. Milakis, M. Snelder, B. Van Arem, B. Van Wee, and G. H. De Almeida Correia, "Development and transport implications of automated vehicles in the Netherlands: Scenarios for 2030 and 2050," *European Journal of Transport and Infrastructure Research*, vol. 17, no. 1, pp. 63–85, 2017.

[2] G. M. Björklund and L. Åberg, "Driver behaviour in intersections: Formal and informal traffic rules," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 8, no. 3, pp. 239–253, 2005.

[3] G. Markkula, R. Madigan, D. Nathanael, E. Portouli, Y. M. Lee, A. Dietrich, J. Billington, A. Schieben, and N. Merat, "Defining interactions : A conceptual framework for understanding interactive behaviour in human and automated road traffic," *Theoretical Issues in Ergonomics Science*, vol. 21, no. 6, p. 29, 2020. [Online]. Available: https://doi.org/10.1080/1463922X.2020.1736686

[4] P. Bazilinskyy, Y. B. Eisma, D. Dodou, and J. C. de Winter, "Risk perception: A study using dashcam videos and participants from different world regions," *Traffic Injury Prevention*, vol. 21, no. 6, pp. 347–353, 2020. [Online]. Available: https://doi.org/10.1080/15389588.2020.1762871

[5] A. Borowsky, D. Shinar, and T. Oron-Gilad, "Age, skill, and hazard perception in driving," *Accident Analysis and Prevention*, vol. 42, no. 4, pp. 1240–1249, 2010. [Online]. Available: http://dx.doi.org/10.1016/j.aap.2010.02.001

[6] W. Vlakveld, "Toetsen en trainen van gevaarherkenning," Stichting Wetenschappelijk Onderzoek Verkeersveiligheid SWOV, Leidschendam, Tech. Rep., 2008.

[7] C. T. Scialfa, D. Borkenhagen, J. Lyon, M. Deschênes, M. Horswill, and M. Wetton, "The effects of driving experience on responses to a static hazard perception test," *Accident Analysis and Prevention*, vol. 45, pp. 547–553, 2012. [Online]. Available: http://dx.doi.org/10.1016/j.aap.2011.09.005

[8] A. F. Williams, "Views of U.S. drivers about driving safety," *Journal of Safety Research*, vol. 34, no. 5, pp. 491–494, 2003.

[9] K. Okamoto, K. Berntorp, and S. Di Cairano, "Driver Intention-based Vehicle Threat Assessment using Random Forests and Particle Filtering," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 13 860–13 865, 2017. [Online]. Available: https://doi.org/10.1016/j.ifacol.2017.08.2231

[10] S. Kolekar, J. de Winter, and D. Abbink, "Human-like driving behaviour emerges from a risk-based driver model," *Nature Communications*, vol. 11, no. 1, 2020. [Online]. Available: http://dx.doi.org/10.1038/s41467-020-18353-4

[11] J. P. Nuñez Velasco, H. Farah, B. van Arem, and M. P. Hagenzieker, "Studying pedestrians' crossing behavior when interacting with automated vehicles using virtual reality," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 66, pp. 1–14, 2019.

[12] L. Malta, C. Miyajima, and K. Takeda, "A study of driver behavior under potential threats in vehicle traffic," *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 2, pp. 201–210, 2009.

[13] S. Breznitz, *Cry Wolf, The Psychology of False Alarms.* New Jersey: Lawrence Erlbaum Associates, Publishers, 1984.

[14] S. Ahmed, M. N. Huda, S. Rajbhandari, C. Saha, M. Elshaw, and S. Kanarachos, "Pedestrian and cyclist detection and intent estimation for autonomous vehicles: A survey," *Applied Sciences (Switzerland)*, vol. 9, no. 11, pp. 1–38, 2019.

[15] P. Dollár, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: A benchmark," *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009*, vol. 2009 IEEE, pp. 304–311, 2009.

[16] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset. The International Journal of Robotics Research," *The International Journal of Robotics Research*, no. October, pp. 1–6, 2013.

[17] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The Cityscapes Dataset for Semantic Urban Scene Understanding," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-Decem, pp. 3213–3223, 2016.

[18] X. Li, F. Flohr, Y. Yang, H. Xiong, M. Braun, S. Pan, K. Li, and D. M. Gavrila, "A new benchmark for vision-based cyclist detection," in *IEEE Intelligent Vehicles Symposium, Proceedings*, vol. 2016-Augus, no. Iv, 2016, pp. 1028–1033.

[19] A. Ess, B. Leibe, K. Schindler, and L. Van Gool, "A mobile vision system for robust multi-person tracking," *26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, 2008.

[20] C. Wojek, S. Walk, and B. Schiele, "Multi-Cue onboard pedestrian detection," *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009*, vol. 2009 IEEE, pp. 794–801, 2009.

[21] K. Gauen, R. Dailey, J. Laiman, Y. Zi, N. Asokan, Y. H. Lu, G. K. Thiruvathukal, M. L. Shyu, and S. C. Chen, "Comparison of visual datasets for machine learning," *Proceedings - 2017 IEEE International Conference on Information Reuse and Integration, IRI 2017*, vol. 2017-Janua, pp. 346–355, 2017.

[22] C. Moran, J. M. Bennett, and P. Prabhakharan, "Road user hazard perception tests: A systematic review of current methodologies," *Accident Analysis and Prevention*, vol. 129, no. February, pp. 309–333, 2019. [Online]. Available: https://doi.org/10.1016/j.aap.2019.05.021

[23] C. M. T. Herdtweck, "Kitti Vision Benchmark," 2021. [Online]. Available: http://www.cvlibs.net/datasets/kitti/raw_data.php

[24] W. Weijermars, "Monitor Verkeersveiligheid 2019. Effectieve maatregelen nodig om het tij te keren," SWOV – Instituut voor Wetenschappelijk Onderzoek Verkeersveiligheid, Tech. Rep., 2019. [Online]. Available: https://www.swov.nl/file/18142/download?token=FVwrxPmz

[25] Central Bureau for Statistics, "Transport en mobiliteit 2016," Centraal Bureau voor de Statistiek, Den Haag, Tech. Rep., 2016. [Online]. Available: https://www.cbs.nl/-/media/_pdf/2016/25/tm2016_web.pdf

[26] SWOV, "SWOV Fact sheet The relation between speed and crashes," SWOV, Tech. Rep. April, 2012.

[27] S. Schmidt and B. Färber, "Pedestrians at the kerb - Recognising the action intentions of humans," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 12, no. 4, pp. 300–310, 2009. [Online]. Available: http://dx.doi.org/10.1016/j.trf.2009.02.003

[28] K. Vogel, "A comparison of headway and time to collision as safety indicators," *Accident Analysis and Prevention*, vol. 35, no. 3, pp. 427–433, 2003.

[29] R. S. Jurecki and T. L. Stańczyk, "Driver reaction time to lateral entering pedestrian in a simulated crash traffic situation," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 27, no. PA, pp. 22–36, 2014.

[30] P. Virtanen, R. Gommers, T. E. T. Oliphant, E. Al., T. E. Oliphant et al., T. E. T. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, I. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, and P. van Mulbregt, "SciPy 1.0: fundamental algorithms for scientific computing in Python," *Nature Methods*, vol. 17, no. 3, pp. 261–272, 3 2020. [Online]. Available: https://doi.org/10.1038/s41592-019-0686-2

[31] A. Yaman, G. Iacca, and F. Caraffini, "A comparison of three differential evolution strategies in terms of early convergence with different population sizes," *AIP Conference Proceedings*, vol. 2070, no. September, 2019.

[32] "Appen Limited," 2021. [Online]. Available: https://appen.com/

[33] J. C. De Winter and D. Dodou, "The driver behaviour questionnaire as a predictor of accidents: A meta-analysis," *Journal of Safety Research*, vol. 41, no. 6, pp. 463–470, 2010. [Online]. Available: http://dx.doi.org/10.1016/j.jsr.2010.10.007

[34] "Heroku." [Online]. Available: https://www.heroku.com/

[35] G. Y. Kanyongo, "The influence of reliability on four rules for determining the number of components to retain," *Journal of Modern Applied Statistical Methods*, vol. 5, no. 2, pp. 332–343, 2006.

[36] M. Brackstone, B. Waterson, and M. McDonald, "Determinants of following headway in congested traffic," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 12, no. 2, pp. 131–142, 2009. [Online]. Available: http://dx.doi.org/10.1016/j.trf.2008.09.003

A Eigenvector matrix

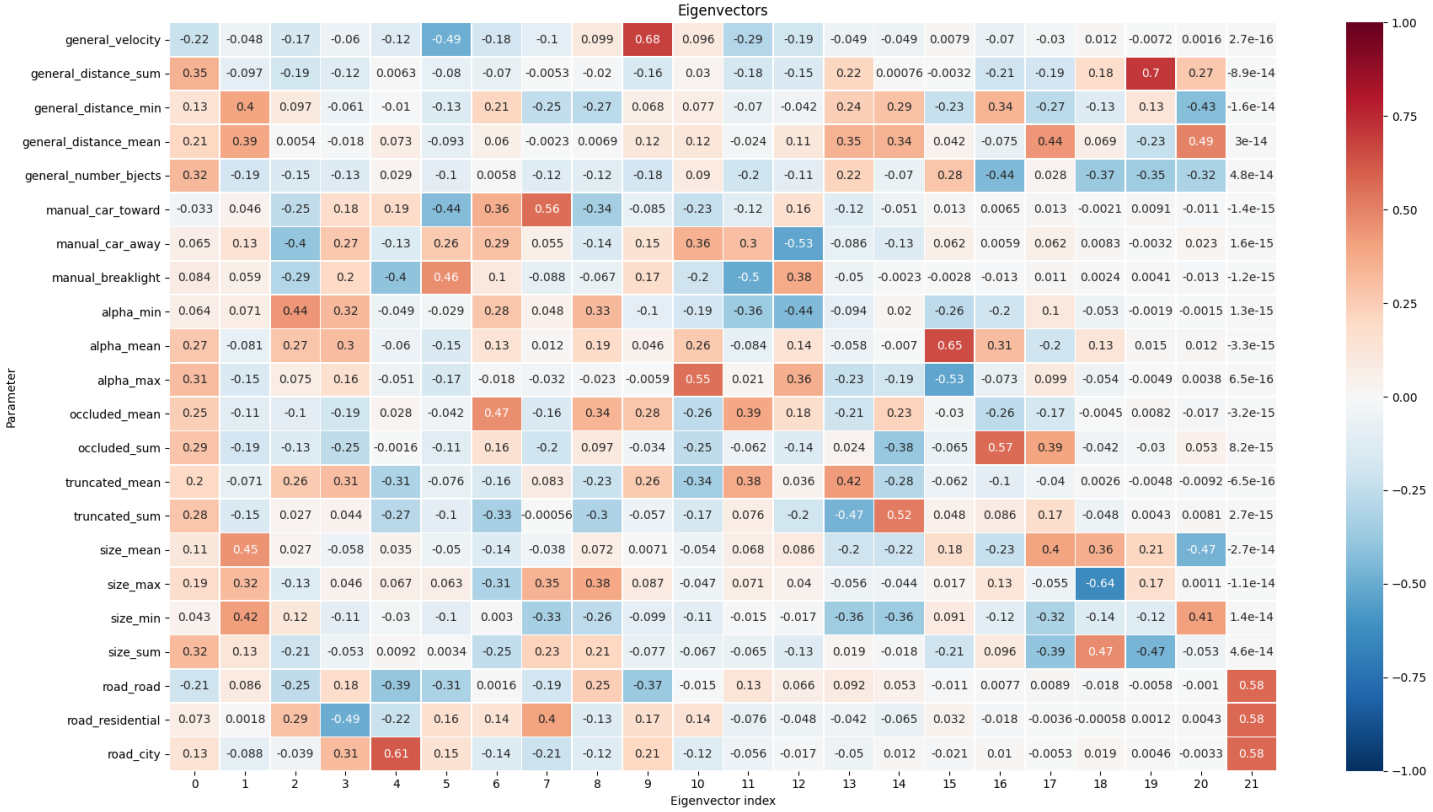

Figure 15: Eigenvector matrix of parameters.

B Correlation matrix before optimisation

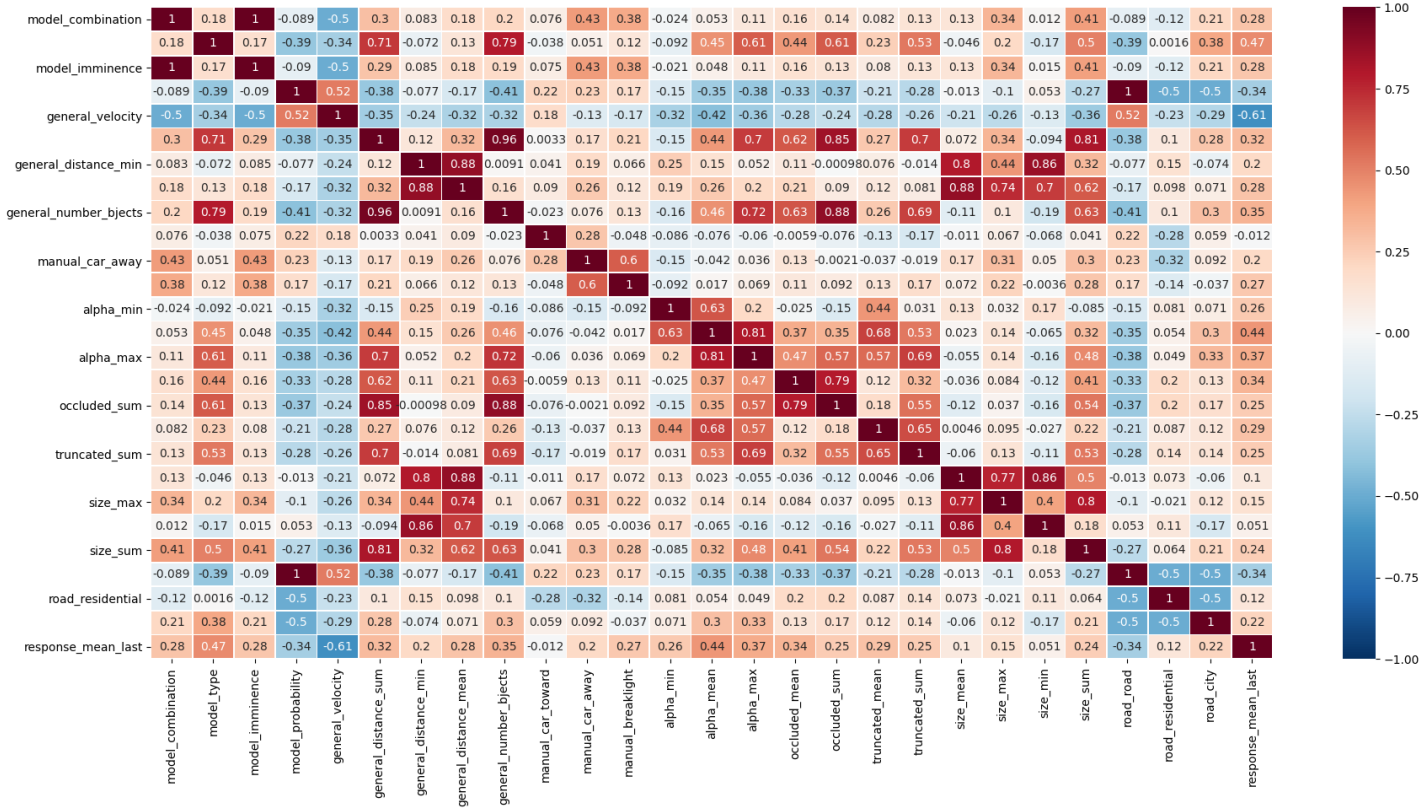

Figure 16: Correlation before optimisation.

C Correlation matrix after optimisation with small bounds

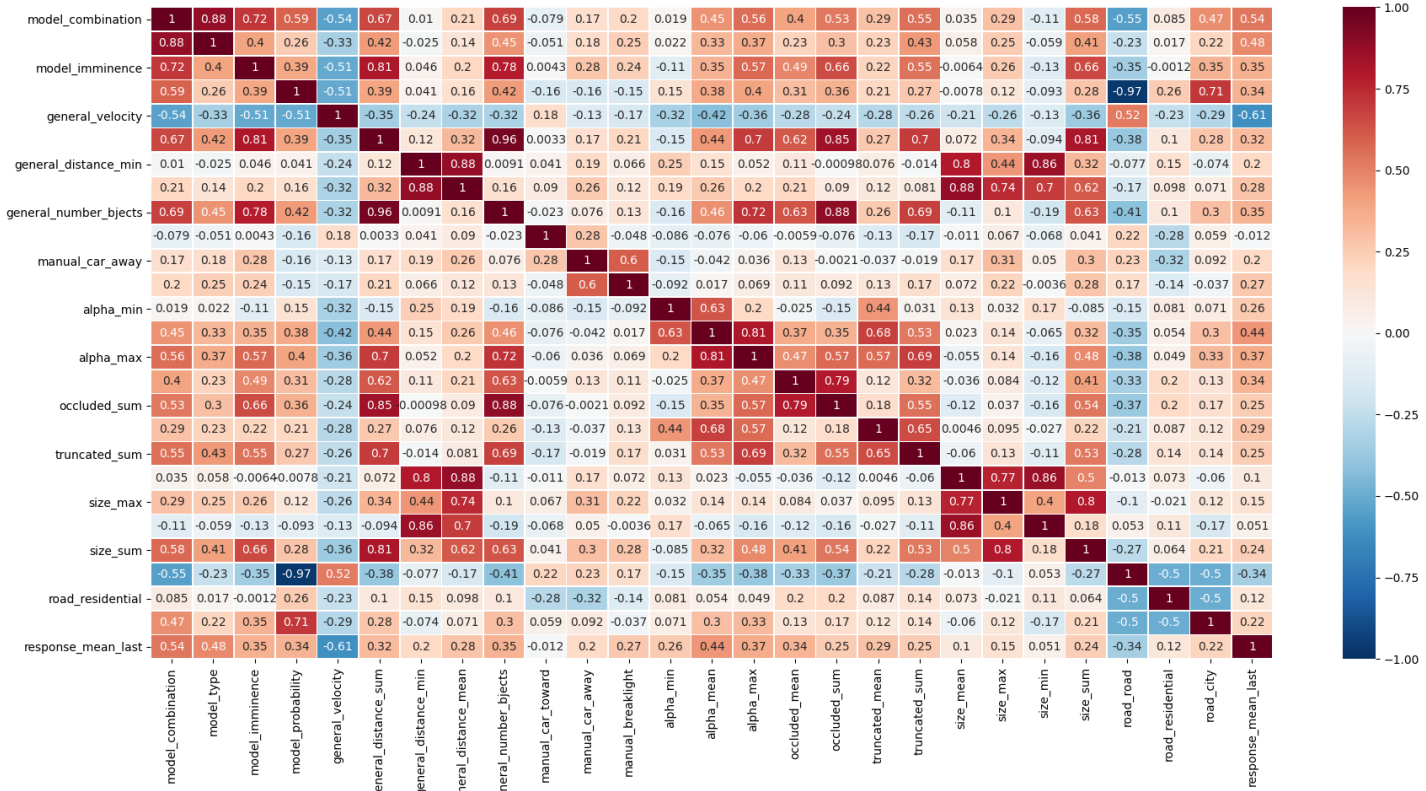

Figure 17: Correlation after optimisation with small bounds.

D Correlation matrix after optimisation with large bounds

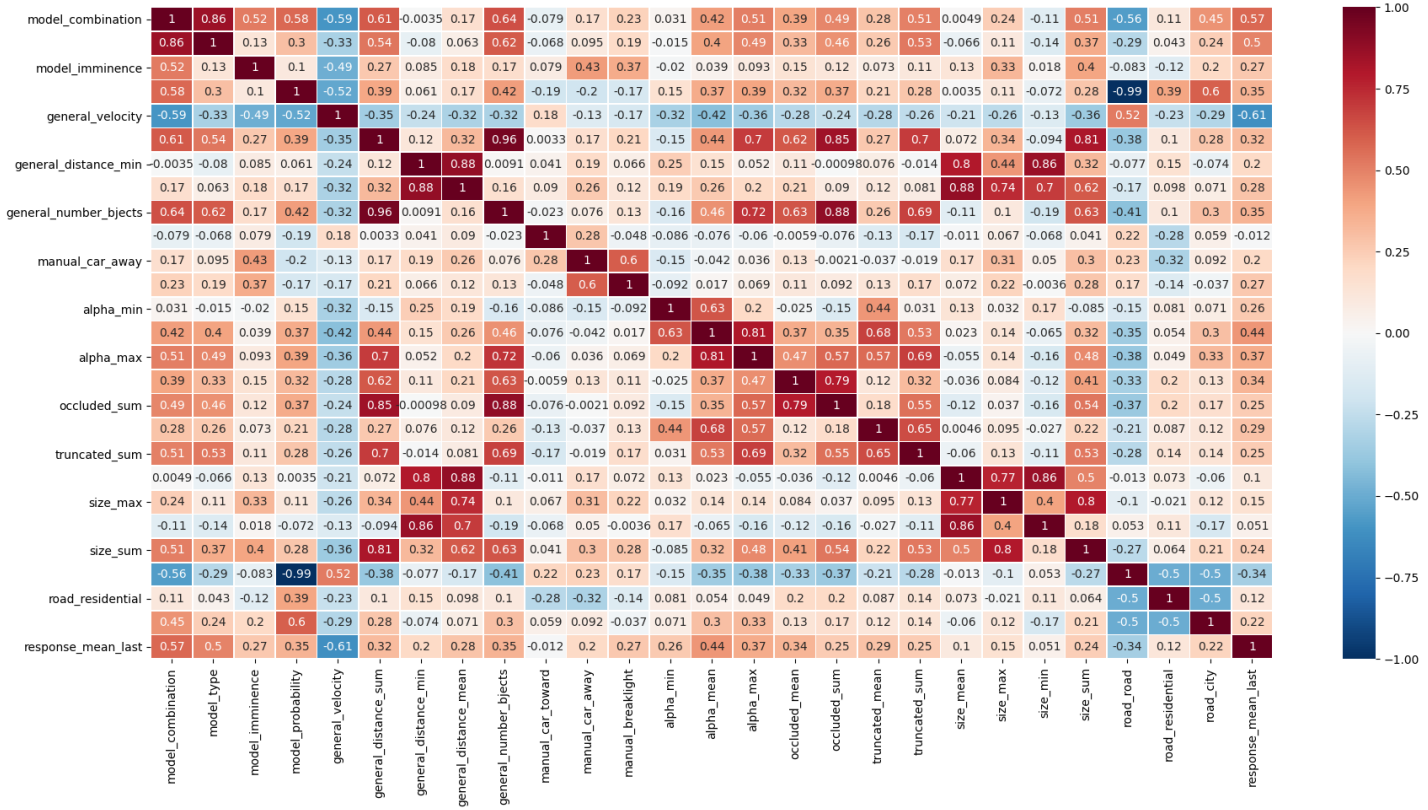

Figure 18: Correlation after optimisation without bounds.

E Online survey

Perceived Risk Of Dash Camera Images

Instructions ▲

You are invited to participate in a research study entitled "Driver threat perception". The study is being conducted by Jim Hoogmoed, master student Mechanical Engineering at the Delft University of Technology, The Netherlands. He is supervised by Dr.ir. Joost de Winter, Ir. Jork Stapel and Dr. Pavlo Bazilinskyy of the Department of Cognitive Robotics, Delft University of Technology, The Netherlands. Contact: j.hoogmoed@student.tudelft.nl (mailto:j.hoogmoed@student.tudelft.nl).

The purpose of this research is 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.

You are free to contact the investigator at the above email address to ask questions about the study. You must be at least 18 years old to participate. The survey will take approximately 25 minutes of your time. In case you participated in a previous survey of one of the researchers of this study, your responses may be combined with the previous survey. The information collected in the survey is anonymous. Participants will not be personally identifiable in any research papers arising from this study. If you agree to participate and understand that your participation is voluntary, then continue. If you would not like to participate, then please close this page. Before the study starts, the images will be preloaded. This may take a few minutes depending on your Internet connection.

General questions

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

- ○ Yes
- ○ No

What is your gender? (required)

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

What is your age? (required)

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

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

- ○ Outdoor, dim light
- ○ Outdoor, bright light
- ○ Other
- ○ I prefer not to respond

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

Which input device are you using now? (required)

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

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

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

What is your primary mode of transportation (required)

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

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

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

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

- ○ 0 km / mi

- ○ 1 - 1,000 km (1 - 621 mi)
- ○ 1,001 - 5,000 km (622 - 3,107 mi)
- ○ 5,001 - 15,000 km (3,108 - 9,321 mi)
- ○ 15,001 - 20,000 km (9,322 - 12,427 mi)
- ○ 20,001 - 25,000 km (12,428 - 15,534 mi)
- ○ 25,001 - 35,000 km (15,535 - 21,748 mi)
- ○ 35,001 - 50,000 km (21,749 - 31,069 mi)
- ○ 50,001 - 100,000 km (31,070 - 62,137 mi)
- ○ More than 100,000 km (more than 62,137 mi)
- ○ I prefer not to respond

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Experiment

You will be asked to leave Appen to participate in the rating task. You will need to open the link below. Do not close this tab. In the end of the experiment you will be given a code to input in the next

question on this tab. Please take a note of the code. Without the code, you will not be able to receive money for your participation. All images will be downloaded before the start of the experiment. It may take a few minutes. Please do not close your browser during that time.

Open **this link (https://risk-dash-crowdsourcing.herokuapp.com/)** to start experiment.

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

Miscellaneous questions

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

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

Test Validators