Article:
Paschalidis, E orcid.org/0000-0001-7648-525X, Choudhury, CF orcid.org/0000-0002-8886-8976 and Hess, S orcid.org/0000-0002-3650-2518 (2018) Modelling the effects of stress on gap-acceptance decisions combining data from driving simulator and physiological sensors. Transportation Research Part F: Traffic Psychology and Behaviour, 59 (Part A). pp. 418-435. ISSN 1369-8478

https://doi.org/10.1016/j.trf.2018.09.019

Modelling the effects of stress on gap-acceptance decisions combining data from driving simulator and physiological sensors

Evangelos Paschalidis, Charisma F. Choudhury, Stephane Hess

Choice Modelling Centre, Institute for Transport Studies, University of Leeds,
36-40 University Rd, Leeds LS2 9JT, UK

Abstract

Driving behaviour is an inherently complex process affected by various factors ranging from network topography, traffic conditions and vehicle features to driver characteristics like age, experience, aggressiveness and emotional state. Among these, the effects of emotional state and stress have received considerable attention in the context of crash analysis and safety research where driving behaviour has been found to be affected by drivers' mental state/stress, cognitive workload and distraction. However, these studies are mostly based on questionnaire surveys and self-reports which can be prone to response bias and reporting/measurement errors. The analyses are also often descriptive in nature. In a parallel stream of research, advances in sensor technologies have made it possible to observe drivers' stress through human physiological responses, e.g. heart rate, electro-dermal activity etc. However, these studies have primarily focused on detecting stress rather than quantifying or modelling its effects on driving decisions. The present paper combines these two approaches in a single framework and investigates the gap-acceptance behaviour of drivers during an intersection crossing, using data collected using a driving simulator. The participants are deliberately subjected to stress induced by time pressure, and their stress levels are measured using two physiological indicators, namely Electrodermal Activity (skin conductance) and heart rate. In addition to statistical analyses, discrete choice models are developed to link the accept-reject choices of a driver with the driver demographics, traffic conditions and stress levels. The results of the models indicate that increased stress levels significantly increase the probabilities of accepting a gap. The improvement in model fit and safety implications derived from model estimates are also discussed. The insights from the results can be used for designing appropriate intervention strategies to improve safety.

Keywords: *gap-acceptance, driving stress, driving simulator, heart rate, Electrodermal Activity*

1. Introduction

Road safety continues to be an important issue with road crashes among the leading causes of death - accounting for more than 1.2 million fatalities and 50 million injuries globally each year (World Health Organization, 2015). Driver behaviour is a factor in over 90% of crashes, with speeding as one of the major contributors (World Health Organization, 2015). Driving behaviour models, which provide mathematical representations of drivers' decisions involving acceleration-deceleration, lane-changing, overtaking, etc., are increasingly being used for evaluation and prediction of road safety parameters and formulating remedial measures (e.g. Farah et al., 2009; Barceló, 2010; Hoogendoorn at el., 2010; Farah & Koutsopoulos, 2014). Reliable driving behaviour models are also critical for accurate prediction of congestion levels in microscopic traffic simulation tools) and analyses of emissions.

Driving decisions are affected by various factors, including network topography, traffic conditions and driver characteristics - which include, among others, demographics, personality traits and emotional state. Existing driving behaviour models address many of these factors, either fully or partially, where the effects of surrounding traffic conditions have received considerable attention (Ossen and Hoogendoorn, 2005; Toledo, 2007; Choudhury, 2007; Marczak et al., 2013 to name a few). However, in most cases, the models do not adequately capture the sophistication of driver behaviour and the causal mechanism behind their observed decisions. In particular, research in other realms, in the context of crash analysis and safety research, has confirmed that driving behaviour is significantly affected by drivers' mental state/mood (e.g. anger) (Garrity and Demick, 2001), cognitive workload (Hoogendoorn et al., 2010), distraction (Young et al., 2007) and fatigue (Thiffault and Bergeron, 2003). Existing work on drivers' stress has mainly focused on the investigation of the relationship between stress and aberrant behaviour and its impact on safety (Ge et al., 2014; Westerman and Haigney, 2000; Hill and Boyle, 2007). However, these studies primarily examined the effects of stress based on self-reported surveys which can be prone to response bias and reporting errors. Indeed, at best, a driver can report an indication of stress levels, but not an objective measure of a physiological state. In addition, many of these studies are largely descriptive rather than relying on detailed modelling work.

In a parallel stream of research, recent advances in sensor technologies have made it possible to measure drivers' stress levels through human physiological responses, e.g. changes in heart rate, electrodermal activity etc. (Healey and Picard, 2005; Ahmed et al., 2015). However, these studies have primarily focused on detecting stress rather than quantifying or modelling its effects on driving decisions in detail.

This paper aims to fill in this research gap by developing gap acceptance models with explicit consideration of the effect of stress on driving behaviour. The gap-acceptance models developed in this research are based on an extensive experimental study in the University of Leeds Driving Simulator (UoLDS) where the drivers have been intentionally subjected to stressful driving conditions caused by time pressure and surrounding traffic conditions. Their choices of accepted gaps have been recorded alongside physiological measurements of stress indicators (Electrodermal Activity and heart rate) and socio-

demographic characteristics (age, gender, experience). A series of gap acceptance models are developed and augmented by continuous physiological measurements.

The remainder of the paper is organised as follows. We first present a review of the literature, followed by the experimental setting and the data analyses. This is followed by a description of the methodological approach of the study. We then present estimation results followed by concluding remarks where insights from the models are discussed.

2. Literature review

2.1. Stress and driving context

'Driver stress' has been defined as a situation that challenges drivers' abilities, reduces their perceived control or threatens their mental/physical health (Gulian et al., 1989). Driver stress can be a consequence of several factors including the direct demands of the driving task, the environmental conditions (e.g. foggy, icy, etc.), network characteristics (e.g. surface characteristics), junction frequency, speed and flow per lane and/or potential secondary tasks, such as use of navigation system, texting, etc. (Hill and Boyle, 2007). Moreover, time urgency and the level of congestion have been identified as two important factors influencing drivers' stress (Hennessy and Wiesenthal, 1999).

There is a substantial body of literature that investigates the effects of stress on driving behaviour. Drivers under stress may be overwhelmed by negative emotions and thus are more likely to get involved in hazardous situations (Ge et al., 2014). Self-reported stress has been linked to aberrant driving behaviour, namely errors and violations (Kontogiannis, 2006). These types of impaired behaviour are related to road crashes and incidents, therefore stress is considered as an issue related to traffic safety (Westerman and Haigney, 2000; Useche et al., 2015, Qu et al., 2014). Moreover, Ge et al. (2014) found that perceived stress is linked to aggressive and risky driving behaviour. Also, Clapp et al. (2011), grouped reactions under stressful situations in three main categories which are the extremely cautious driving behaviour, aberrant behaviour and aggressive (or hostile) behaviour. The aforementioned findings provide compelling evidence regarding the effects of stress on driving, however, they are based on self-reported survey results and therefore prone to response bias and reporting/measurement errors.

An alternative, and potentially more reliable, approach to detect drivers' level of stress and study its effects, is through its implications on human physiology. Recent advances in sensor technologies and affective computing have made it possible to measure drivers' stress levels through physiological responses, e.g. changes in heart rate, Electrodermal Activity (EDA), blood volume pulse, etc. There are several existing studies related to driving stress that use this type of data (some examples Healey and Picard, 2005; Singh and Queyam, 2013; Rigas et al., 2012). However, the aforementioned studies mostly focused on detecting stress rather than investigating its effects on observed driving behaviour.

Two of the most widespread physiological indicators - also used in the present study - are heart rate and Electrodermal Activity (EDA). Heart rate represents the observed heartbeats

per minute. Lower heart rate is generally linked to a relaxed state while it increases under the presence of emotional stimuli or mental effort (Katsis et al., 2011). EDA is related to the sweat gland activity and it is an indicator that increases or decreases proportionally to stress effort (Katsis et al., 2011). EDA is composed of two different parts, namely the skin conductance level (SCL – tonic part) and skin conductance response (SCR – phasic part). While SCL is slowly varying and related to individual characteristics, SCRs are expressed as a sudden and fast increase of skin conductance owing to the presence of a specific stimulus and thus have been linked to acute stress. SCRs are identified if the increase in skin conductance activity exceeds specific critical values.

Before proceeding, let us just expand on the argument of why such physiological measurements are superior to self-reported measures. The two most apparent issues with self-reported data are perception bias and measurement error. For the former, a respondent to a questionnaire may perceive to be more or less stressed than he/she actually is, and this can be amplified in the case of recall surveys. For the latter, it is difficult for a survey respondent to quantify the level of stress in an objective manner. An additional reason, which is mentioned less often, is that of strategic bias. A respondent in a survey may purposefully overstate or understate his/her actual stress levels for example to make an experienced situation seem more stressful or play down the effect of his/her own mental state. None of these issues should in theory arise with physiological measurements as they are driven by subconscious factors that cannot be easily biased by the respondent and are also measured objectively.

2.2 Gap-acceptance behaviour and models

Driving behaviour models primarily include car-following, lane-change and gap-acceptance (Toledo, 2007). The latter of the aforementioned concepts focuses on two different aspects; the decision of drivers to change lane and the attempt of a turning or crossing manoeuvre at an intersection. In the literature, several methodological approaches have been developed in order to predict the intersection crossing decisions of drivers. This type of gap acceptance behaviour is of prime importance when studying issues such as network capacity, delays and road safety (Ashton, 1971; Fitzpatrick, 1991). The majority of these methodologies are based on the critical gap concept, which is defined as the minimum time gap in the priority stream which a driver moving on the minor road is willing to accept in order to cross through the conflict zone. According to Brilon et al., (1999), there are at least 20-30 different methods related to gap-acceptance decisions. Some of the most cited are the Raff method (Raff and Hart, 1950), the Greenshields method (Greenshields et al., 1946), the lag method (see Brilon et al., 1999), the logit method (Maze, 1981) - which is a method based on traditional choice modelling techniques (see Ben-Akiva and Lerman, 1985), the Ashworth's method (Ashworth, 1969), and the maximum likelihood method (Miller and Pretty, 1968). The main limitations regarding some of the existing methodologies in the context of unsignalised intersections are the assumptions of consistency and homogeneity (Bottom and Ashworth, 1978; Pollatschek et al., 2002). The former indicates that a driver, in all similar situations, would have a specific critical gap value t_c and accept all gaps with a value greater than this (and reject the rest). Based on this assumption, a driver waiting to cross a junction, cannot reject a specific gap and later accept a shorter one. The assumption of consistency is not however accurate since e.g. risk

tolerance of an individual might change during waiting time leading to acceptance of a shorter gap compared to the ones rejected earlier (Pollatschek et al., 2002). Moreover, the various t_c values of different consistent drivers are treated as a random variable that follows a specific distribution $\varphi(t_c)$ and cumulative distribution $\Phi(t_c)$ (Brilon et al., 1999). Sub-groups of drivers are assumed to follow the same density and cumulative distribution functions resulting within-group homogeneity of the driver population.

The assumptions of homogeneity and consistency of gap-acceptance methodologies raise limitations in the representation of drivers' behaviour since they both ignore their sophisticated decision-making process. For instance, critical gap varies among and within drivers, in different situations, and should be treated as a random variable (Guo et al., 2014). The drawbacks imposed by these assumptions have been relaxed in gap-acceptance models developed in the context of lane-changing, where critical gaps are assumed to follow statistical distributions with means being functions of influencing variables like speed of the lead and lag vehicles (e.g. Ahmed 1999, Toledo 2003, Choudhury 2007). These models are also extended to incorporate the effect of driver demographics (age, gender) and driving style (e.g. Farah et al. 2009). Another competing approach is to model the gap accept-reject decisions based on 'Utility maximization theory' – Logit models for example. In Logit models, the probability of accepting or rejecting a gap is a function of different variables (e.g. gap size, the speed of the approaching vehicles, waiting time, etc.) and captures the trade-off among different influencing factors (e.g. Amin and Maurya, 2015).

A review of the gap-acceptance literature showed that drivers' behaviour is influenced by various factors. Most of the variables are related to traffic conditions such as gap size (Bottom & Ashworth, 1978; Nabaee et al., 2011), waiting time in the queue (Pollatschek et al., 2002) or at the stop line (Mahmassani and Sheffi, 1981) and the queue behind the driver while waiting at the stop line (Nabaee et al., 2011; Tupper et al., 2011). Apart from the aforementioned factors, Bottom & Ashworth (1978) mention that inter-individual variance is worth being investigated in terms of variables as extroversion (personality), age, annual mileage and vehicle type.

Despite the advances in gap-acceptance model structures, the full range of variables influencing the decisions of the drivers has not yet been fully investigated. Some of the aspects which are not yet addressed include drivers' strategies when deciding to cross an intersection or not, the motivation behind an observed "inconsistent" action and finally the effects of individual traits and characteristics (e.g. personality, attitudes, state of mind, level of stress etc.). The aim of the present study is to provide an extended gap-acceptance framework, through the development of a model that accounts for variables related to driver's individual characteristics, with explicit consideration of drivers' acute stress levels, and contribute to filling in this gap of driving behaviour modelling research.

3. Data collection

3.1 Driving simulator experiment

The data used in this research is based on primary data collected as part of a comprehensive driving simulator study (Next Generation Driving Behaviour Models – NG-DBM) for

investigating the effect of stress in different driving decisions (e.g. acceleration-deceleration, overtaking, red light violation, gap acceptance, etc.). The experiments have been conducted using the University of Leeds Driving Simulator (UoLDS). The UoLDS (Figure 1) is a high fidelity, dynamic simulator. The vehicle cab is a 2005 Jaguar S-type with all driver controls available and fully operational. This includes the steering wheel and braking pedal, and there is also a fully operational dashboard. The vehicle is positioned in a 4m diameter spherical projection dome. The dome provides fully textured 3-D graphical scene with a horizontal field of view of 250º and 45º vertical and it is placed on an 8 degrees of freedom motion system. The model of vehicle dynamics has been extensively validated to capture accurate vehicle behaviour on high-friction surfaces (Markkula et al., 2018). The raw data output consists of observations of 60Hz frequency. The relative validity of UoLDS has been confirmed in several studies (e.g. Jamson et al., 2010; Markkula et al., 2018). While driving simulator data, given its 'experimental' flavour, has the risk to be prone to behavioural incongruence, it offers the flexibility to fully control the surrounding traffic and driving contexts (e.g. inducing time pressure and stressful scenarios) which are crucial for this particular study.

Figure 1: The University of Leeds Driving Simulator
[sources: University of Leeds, University of Leeds Driving Simulator]

The full data collection process involved around 90 minutes of total driving in the simulator for each individual. Participants initially had a short briefing session about the simulator and its operation followed by a practice session of approximately 15 minutes duration to familiarise themselves with the simulated environment and vehicle dynamics (i.e. motion system). For safety reasons, participants were accompanied by a researcher during the practice run, positioned in the back seat. After the practice session, participants started the main driving sessions, composed of two different environments, using an urban setting and a motorway setting, with a short break in between.

The urban setting was composed by several tasks. These included an encounter with a slow-moving lead vehicle that participants could decide to overtake or not, a traffic light with a red indication of long duration that aimed to cause frustration, an amber dilemma scenario where participants could decide to accelerate or brake and the gap-acceptance scenario presented in the current analysis. These scenarios were repeated twice (without and under the presence of time pressure) while in the end there was also a right-turn manoeuvre scenario which was the last task of the urban setting. Within an effort to minimize any potential residual effects from the previous tasks, some straight road segments without any critical events were included, in between the main tasks. The average duration of these dummy segments was 2-3 minutes and participants did not meet any traffic in these, however, at the second half of the urban setting they were deliberately subjected to time

pressure. The latter needs some more explanation. As mentioned above, the majority of the scenarios had two variants - one without and one with time pressure. Before each of the two main driving simulator settings, participants were instructed that they had to reach the destination within 35 minutes and they could see an emoji placed on the dashboard (Figure 2) denoting their performance with respect to time. Participants were told that the emoji displayed to them was determined based on expected arrival time which is computed and constantly updated using a sophisticated algorithm running in the background and uses variables such as current speed, speed limit, distance to the end, an average estimated delay that will be caused by the events ahead etc. as inputs. This was then used to determine which of the three emoji to show. Participants were instructed that the green state would indicate they were doing well, in terms of time, while the red would mean that they were late. The intermediate amber emoji meant that they were marginally fine in terms of time. That is, they will receive a red emoji if they have further delay in the remaining driving tasks. An amber state was introduced to make the shift from green to red emoji (and vice versa) more convincing to the participants. In reality, the state of the time pressure emoji was not related to their actual performance but was pre-decided in order to induce time pressure in specific road segments. It should be mentioned that the amber was always shown before/after the critical sections (e.g. in straight segments) as opposed to near intersections. Therefore, the data used for gap-acceptance model development only include red and green phases. It may be noted that the choice of 3 different emoji to indicate time pressure, was preferred to a conventional countdown timer since it would be easier to manipulate. In order to increase the likelihood that participants would consider time pressure indications, they were instructed that a penalty would be imposed on the monetary reward they received for their participation in case they were late at the end of a scenario (red emoji). Again, this was never the case since both main scenarios were programmed to end in the amber time pressure state.

Figure 2: Time pressure indications

Drivers' physiological data, across the whole experiment, was collected using the Empatica E4 wristband which is a non-intrusive device that provides information about heart rate (HR), Electrodermal Activity (EDA), blood volume pulse (BVP) and temperature (TEMP). Each of the physiological indicators was collected with a different frequency, depending on the attributes of the wristband. EDA and temperature have a 4Hz frequency, blood volume pulse 64Hz and heart rate 1Hz. The device can be automatically synchronised with the clock of any computer when plugged in.

3.2 The gap-acceptance task

In the present study, the gap-acceptance task was presented twice, as a part of the urban driving scenario. Drivers faced the first gap-acceptance task without time pressure (green

emoji) followed by the same scenario with time pressure (red emoji). The scenario itself consisted of two groups of vehicles. At first, six blocking vehicles were shown to participants, moving at short headway distances. These vehicles were used to force drivers to stop before the main gap acceptance task. This first group of vehicles was followed by eleven vehicles that created 10 gaps. The gaps had an increasing trend in general. The increasing trend of gaps was chosen in order to secure that drivers would not face a large gap at the beginning of the scenario and miss information related to their willingness to accept a shorter one. However, to increase realism, some shorter gaps were also introduced in between (as 3rd, 5th, 7th and 8th gaps). The full set of available gaps were identical for both intersections and across participants. For each gap-acceptance task, the drivers could choose to accept the available gap and cross or reject the immediate gap and wait for a better one or even reject them all (i.e. wait till all 11 vehicles had crossed). The drivers, however, had no a priori knowledge regarding the number of the oncoming vehicles or the waiting time required. For the sake of simplicity, it was decided to constrain the gap-acceptance scenario by developing a case where cars were shown only coming from the left side of the driver. It may be noted that the time pressure was always applied at the second intersection albeit the fact that there might be confounding with learning[1] and fatigue effects. The main reason for this design was related to drivers' physiology, since we aimed to minimise the risk of increasing their responses at the beginning of the driving task by inducing additional stressors (e.g. time pressure) that would potentially influence and prevent them from returning to the baseline levels. Also, it would be more realistic for the participants to receive a red face indication closer to the end of the driving task, rather than during the first part. A general outline of the gap-acceptance scenario setting is illustrated in Figure 3, while the presented gap sizes are shown in Table 1.

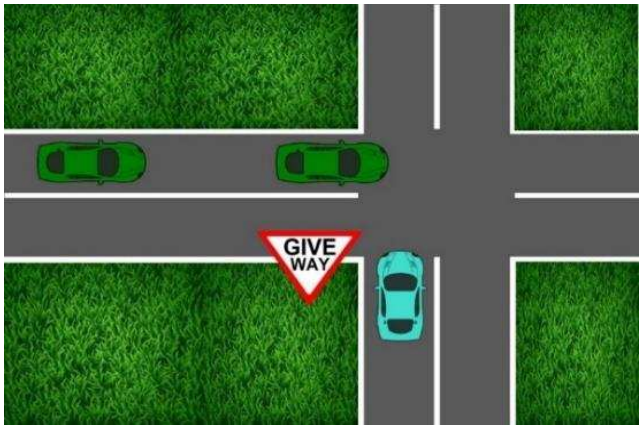

Figure 3: Illustration of the intersection

Table 1: The available gaps and gaps' sizes

Gap ID	1	2	3	4	5	6	7	8	9	10
Gap size (s)	2.8	3.45	3.4	4.4	4	5.4	5	4.7	6	6.8

[1] Since the two scenarios occurred with a time gap of approximately 15 minutes in between where the drivers had to tackle other difficult situations, the learning effect is not expected to be significant.

3.3. Exploratory analysis

3.3.1 Sample analysis

The sample of the current analysis consists of 41 (22 male, 19 female) staff members or students at the University of Leeds, holding a valid driving licence, that successfully completed the urban task. Three participants were removed from the analysis, since they reported motion sickness during the practice session while also an additional participant was removed because the wristband device failed to collect physiological data. The mean age of participants is approximately 34 years and the corresponding standard deviation is 11 years. Almost half of the participants stated that they are driving on a daily basis. The average driving experience of participants is almost 14 years. Regarding accident involvement, 6 participants have reported involvement in minor accidents while 4 have reported involvement in serious accidents. It is worth mentioning that a serious (or major) accident is defined as one where at least one person required medical treatment and/or there was property damage above £500. Finally, 7 participants stated that they had at least once received a ticket penalty for speeding behaviour. The descriptive statistics of the sample are also outlined in Table 2.

Table 2: Descriptive statistics of the sample

Variable	Intervals	Frequency	%	mean	std. dev.	min	max
Gender	Female	19	0.46	-	-	-	-
	Male	22	0.54	-	-	-	-
Age	-	-	-	34.39	10.86	19	57
Driving experience	-	-	-	13.63	11.48	1	39
Frequency of driving	Everyday	21	0.51	-	-	-	-
	2-3 times/week	12	0.29	-	-	-	-
	Once/ week	4	0.10	-	-	-	-
	Less often	4	0.10	-	-	-	-
Minor accident involvement	No	35	0.85	-	-	-	-
	Yes	6	0.15	-	-	-	-
Major accident involvement	No	37	0.90	-	-	-	-
	Yes	4	0.10	-	-	-	-
Ticket for speeding	No	34	0.83	-	-	-	-
	Yes	7	0.17	-	-	-	-

3.3.2 Gap-acceptance task analysis

Before the development of the model, participants' gap-acceptance behaviour has been examined with respect to the effects of time pressure. Table 3 presents the accepted gaps of each individual, and their respective size (a value n/a is given if no gap is accepted). A similar illustration is also provided in Figure 4. It should be mentioned that 12 out of 41 participants did not accept any of the gaps presented to them (i.e. waited for all vehicles to pass), in both cases. On the other hand, six participants accepted a gap only under the time pressure conditions while they had not done so without time pressure. The remaining 23 participants accepted a gap at both intersections. The latter group of participants always accepted the same gap in the second run or a gap that was shown earlier, compared to the one accepted without the external stressor. To further investigate this outcome, a paired samples t-test is applied to compare the significance of the difference of the accepted gap sizes at the two intersections. Given the small sample size, this difference has been also

Table 3: Accepted gap(s) of each participant at the two intersections

ID		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
First intersection (without time pressure)	Gap ID	8	9	11	11	5	11	11	6	6	9	6	9	11	11	5	5	11	11	6	6	9
	Gap size (s)	4.7	6	n/a	n/a	4	n/a	n/a	5.4	5.4	6	5.4	6	n/a	n/a	4	4	n/a	n/a	5.4	5.4	6
Second intersection (under time pressure)	Gap ID	1	5	9	11	4	11	11	4	6	9	4	6	11	7	1	1	11	11	4	4	9
	Gap size (s)	2.8	4	6	n/a	4.4	n/a	n/a	4.4	5.4	6	4.4	5.4	n/a	5	2.8	2.8	n/a	n/a	4.4	4.4	6

ID		22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41
First intersection (without time pressure)	Gap ID	8	6	11	6	11	6	11	11	11	8	9	11	5	9	9	11	11	11	5	11
	Gap size (s)	4.7	5.4	n/a	5.4	n/a	5.4	n/a	n/a	n/a	4.7	6	n/a	4	6	6	n/a	n/a	n/a	4	n/a
Second intersection (under time pressure)	Gap ID	4	3	11	6	6	6	11	11	11	4	9	11	4	5	6	9	9	11	4	6
	Gap size (s)	4.4	3.4	n/a	5.4	5.4	5.4	n/a	n/a	n/a	4.4	6	n/a	4.4	4	5.4	6	6	n/a	4.4	5.4

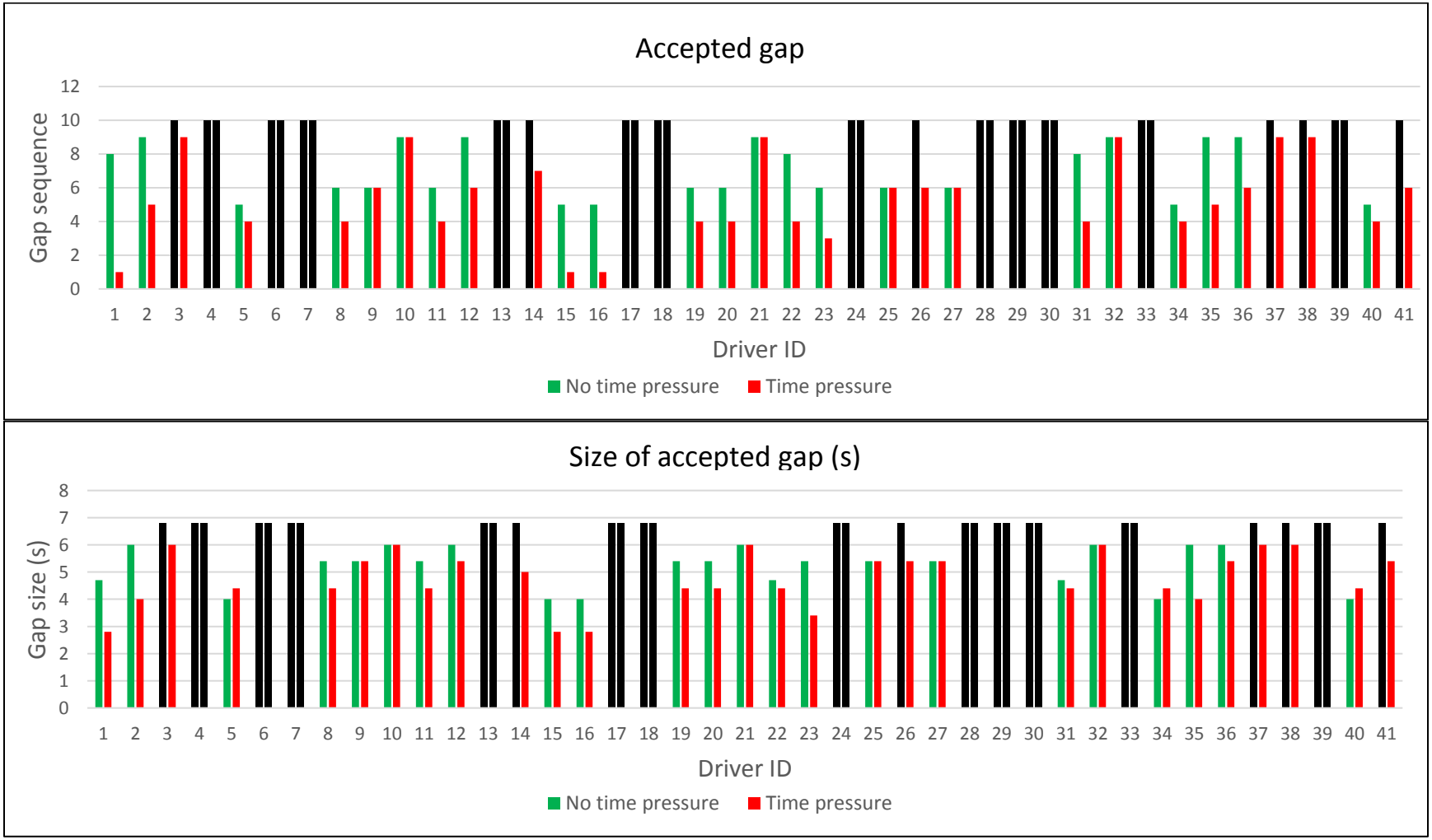

Figure 4: Accepted gaps and sizes without and with time pressure

investigated with the non-parametric Wilcoxon test (De Winter 2013). The results (Table 4) show that the mean size of the accepted gaps is smaller at the second intersection, and this difference is statistically significant. As mentioned in the data collection section, since the participants faced a series of additional tasks involving at least 15min of driving in between the two intersections, the learning effect is not likely to be a major influencing factor behind these choices. We, therefore, conclude that time pressure had a major influence on acceptance of smaller gaps which we further test empirically in Section 5. The mean values in Table 4 are smaller than some reported in the existing literature (e.g. Bottom & Ashworth, 1978; Fitzpatrick, 1991) however, they are very close to the median values reported by Ashton (1971) and Amin and Maurya (2015). It may be noted that given the simulated nature of our experiment and the scope to show a limited number of gaps to each participant, the presented gaps were on the shorter range on purpose. Otherwise, there would have been risk of missing the minimum acceptable gap.

Table 4: Results of the paired samples t-test and Wilcoxon test

Descriptives				**Paired samples t-test**						
	Mean	**SD**	**SE**	**T**	**df**	**P**	**Mean Difference**	**SE Difference**	**95% CI for Mean Difference**	
									Lower	**Upper**
First intersection (no time pressure)	5.191	0.768	0.16							
Second intersection (under time pressure)	4.558	0.978	0.204	3.752	22	0.001	0.633	0.169	0.283	0.983
Wilcoxon test p-value: 0.002										

Furthermore, with reference to Table 3, under the time pressure conditions, three of the participants accepted the first gap they faced. These drivers did not actually behave as expected during the task (stop at the intersection and wait for a gap, or not, to cross) but drove through the streaming of oncoming vehicles without stopping. This indicates that external stressors could increase risk-taking – however, such extreme behaviour may not be frequently observed in real life. Moreover, it is worth mentioning that the 10^{th} gap was never accepted in the current experiment, although it is the largest one in terms of headway size. This behaviour maybe shows anticipation effects in gap-acceptance behaviour; drivers that wait until the last available gap also prefer to wait the additional time need until being able to cross when the intersection is clear rather than engaging in crossing under the presence of oncoming traffic. As mentioned above, almost one-third of participants follow this behaviour, without being influenced by time pressure in the second task.

4. Methodology

4.1 The gap acceptance model

The gap-acceptance approach of the current paper has been formulated as a binary choice model, where each gap is considered as a different accept/reject decision. This approach is a modification of the Logit method mentioned in the literature section. The model assumes that the probability of accepting a gap increases with the increase in the utility. The utility associated with a particular gap is a function of the attributes of the gap (e.g. gap size, order, etc.), characteristics of the driver (e.g. socio-demographics) and their state. The utility U_{nt}

associated with the decision of a driver *n* to accept/reject a gap *t* can therefore be expressed as follows:

$$U_{nt} = \beta X_{nt} + \gamma Z_n + \theta W_{nt} + \alpha \nu_n + \varepsilon_{nt} \quad (1)$$

where X_{nt} is a vector of gap-specific variables, Z_n are individual-specific and situation-independent variables (e.g. socio-demographics), W_{nt} is a vector of physiological variables that are used to capture drivers' mental state, ν_n represents the effect of unobserved variables that vary across individual drivers but is same for a specific driver (referred as individual specific error term), and ε_{nt} is the random error term (assumed to be independent and identically distributed). Finally, β, γ, θ and α are vectors of parameters to be estimated.

Following the aforementioned assumptions, the probability of gap-acceptance conditional on individual specific error term is defined as:

$$P_{nt}^{GA}|\nu = \frac{e^{(\beta X_{nt} + \gamma Z_n + \theta W_{nt} + \alpha \nu_n)}}{1 + e^{(\beta X_{nt} + \gamma Z_n + \theta W_{nt} + \alpha \nu_n)}} \quad (2)$$

If the observed choice of a driver to accept a gap is set as Y_{nt}=1, the conditional full probability of an observed driver's decision can be expressed, as shown in Equation 3:

$$P_{nt}|\nu = (P_{nt}^{GA}|\nu)^{Y_{nt}}(1 - P_{nt}^{GA}|\nu)^{1-Y_{nt}} \quad (3)$$

The conditional probability of a sequence of T_n observed decisions of the same driver takes the form indicated by Equation 4:

$$P_n|\nu = \prod_{t=1}^{T_n}(P_{nt}|\nu) \quad (4)$$

The unconditional joint probability of the observations of a given driver can be expressed as follows:

$$P_n = \int_{-\infty}^{+\infty}(P_n|\nu)\varphi(\nu)d\nu \quad (5)$$

where a $\varphi(\nu)$ is the probability density function of the individual specific error term assumed to have a standard normal distribution. The model parameters are jointly estimated using the Simulated Maximum Likelihood approach using 1000 Halton draws (Halton 1960). The model has been specified and estimated in R based on the code framework provided by the Choice Modelling Centre, University of Leeds.

4.2. Physiological data analysis

The model described in the previous section has been augmented by continuous

physiological measurements. These observations have been used as direct explanatory variables, in order to investigate whether the gap-acceptance model would be more behaviourally representative when stress has been included. Two different responses have been considered, namely, heart rate and Electrodermal Activity (EDA). Before turning to the actual implementation, it is worth briefly discussing our use of these measures as direct explanators. Recent work in choice modelling (Abou-Zeid and Ben-Akiva, 2014) has focussed on the use of hybrid choice models to incorporate additional indicators of heterogeneity such as answers to attitudinal questions. This type of approach is not critical in our case as the physiological measures are direct measures of physiological states and should thus not be affected by the same concerns of measurement error.

The physiological variables have been initially processed and transformed before their incorporation in the model. Transformation or standardisation of physiological variables is a common practice in relevant research (e.g. Zhai and Barreto, 2006; Singh et al., 2013; Kalimeri & Saitis,2016), within an effort to reduce the inter-individual differences in physiological responses, while it has also been found to improve the distinction among the various physiological states (Ben□Shakhar, 1985). In the current approach, each gap is considered as a different discrete stimulus, rather than assuming the whole sequence as a single continuous stimulus (differences between the two approaches are explained in Cacioppo et al. (2007)). Thus, physiological responses used in the model have been calculated with reference to the initiation of each gap (i.e. when the lead vehicle associated with the gap reaches the beginning of the intersection).

Instead of using the raw observations, the heart rate data have been normalised at the individual level, applying a z-score transformation $\left(\frac{x-\mu}{\sigma}\right)$, where x is a heart rate observation, μ is the heart rate mean value across the whole urban task and σ is its standard deviation (Picard et al., 2001; Healey and Picard, 2005; Maaoui and Pruski, 2010). The normalized heart rate in the beginning of each gap is then considered as a variable in the model.

The EDA observations have been processed using the Matlab package Ledalab (Karenbach, 2005). The skin conductance responses (SCRs) have been obtained applying trough-to-peak analysis, where the amplitude of a response is calculated as the difference in the EDA values between a peak in the signal and its preceding trough (Benedek and Kaernbach, 2010). The amplitude is then considered as an explanatory variable in the models. The EDA analysis is based on event-related response activation; each gap has been considered as a different stimulus. The initiation of each gap has been used as the starting point and responses are detected 1-4s after that moment. Moreover, since –we are interested to capture the stress-level at the beginning of a gap (when the lead vehicle corresponding to the gap reaches the intersection), the amplitudes corresponding to the immediately preceding gap has been used as an explanatory variable. An example of SCRs analysis is illustrated in Figure 5. Following literature indications (e.g. Sano et al., 2014), a critical value equal to 0.01μS is selected as a minimum critical SCRs. Moreover, each significant amplitude (above 0.01μS) has been divided by the maximum observed SCR amplitude, during the simulator experiment, to minimise the effects of individual differences (Lykken, 1972).

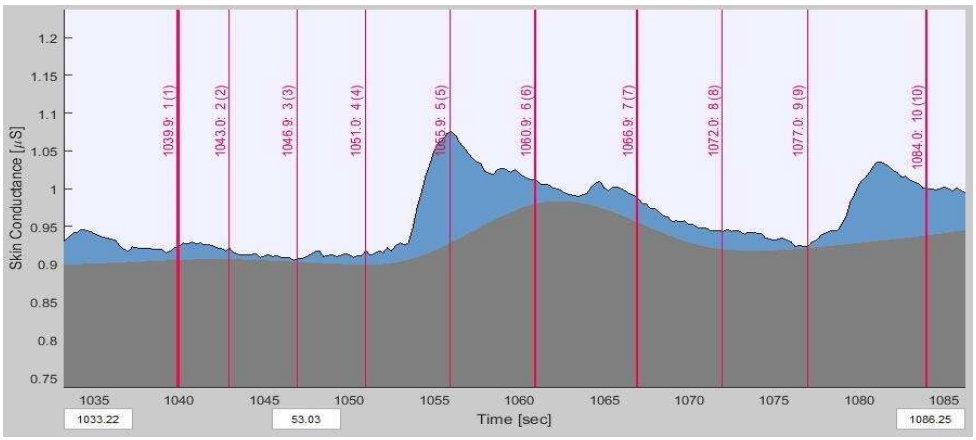

Figure 5: Example of SCRs extraction

5. Gap-acceptance model

5.1 Parameter estimation results and interpretation

A series of gap-acceptance models have been estimated based on the methodology presented in section 4.1. The first model includes only traffic-related variables, while the socio-demographics, time pressure dummy, and the physiological observations are eventually added. Thus, each new model includes all the previous variables plus one or more new ones. The aim of this approach is to compare model fit and investigate the incremental improvement (if any) of adding a specific group of variables. Four different models have been estimated in total, as follows:

- Model 1: Traffic-related variables only
- Model 2: Socio-demographics variables included
- Model 3: Time pressure considered
- Model 4: Physiological variables included

The results of all four models are presented in Table 5. All parameter estimates are significant at 95% level (|t-ratio|>1.96).

With reference to Table 5, gap size, speed, position, skin conductance response (SCR) and heart rate are continuous variables explained in the next paragraphs. Moreover, a series of dummy variables have been included in the models.

Model 1: Traffic-related variables only

The first model includes the gap size (in seconds), the position of the vehicle during the waiting time, vehicle speed when arriving at the intersection area, a dummy variable indicating whether there is another gap following, or not, and the standard normal disturbance term (Model 1) as explanatory variables.

As expected, gap size has a positive effect on gap acceptance behaviour showing that drivers' probability to accept a gap increases with its size.

Table 5: Gap-acceptance models' parameter estimates

Variable	Traffic related variables only model Model 1		Socio-demographics included model Model 2		Time pressure included model Model 3		Physiological observations included model Model 4	
	Estimate	t-ratio	estimate	t-ratio	estimate	t-ratio	estimate	t-ratio
Constant	-24.39	-3.74	-25.92	-3.92	-33.56	-5.14	-59.05	-7.32
Gap size	1.14	2.50	1.28	2.87	1.99	3.66	2.87	3.23
Last gap dummy	-9.01	-15.16	-9.17	-14.78	-10.24	-14.69	-13.57	-7.31
Speed (first gap)	2.09	3.77	1.98	3.90	2.18	4.54	3.96	6.10
Position	2.92	3.40	2.77	3.48	3.15	3.92	6.13	5.04
α^{acc}	4.59	4.38	3.47	3.92	4.70	4.48	8.89	4.93
Age>45 dummy	-	-	-5.87	-4.16	-7.89	-4.29	-11.38	-3.79
Regular driver dummy	-	-	5.00	3.05	6.50	3.18	10.08	3.27
Time pressure dummy	-	-	-	-	2.45	2.96	3.75	2.94
Skin conductance response	-	-	-	-	-	-	11.90	2.62
Heart rate	-	-	-	-	-	-	2.40	2.13
LL0		-426.29		-426.29		-426.29		-426.29
LL		-83.53		-75.82		-71.61		-64.90
ρ^2		0.80		0.82		0.83		0.85
adjusted ρ^2		0.79		0.80		0.81		0.82
observations		615		615		615		615

Vehicle position is a variable that captures a vehicle's position at the intersection area (the value zero denoting the start of the intersection area) with an increase in value as the vehicle moves forward. If a participant has been outside of the intersection area during the task (it is the case for some participants during the first shown gap), the variable could also take negative values. The inclusion of this variable attempts to capture drivers' behaviour to better position themselves and increase the likelihood of accepting the next available gap. This variable was considered in the model as, during data collection, a proportion of drivers was observed to slowly move their car forward during the period they were waiting for an acceptable gap. As expected, the effect of this variable is positively related to the gap-acceptance probability and drivers are more likely to accept a gap the closer to or further inside the intersection their vehicle is.

The variable vehicle speed is considered in the utility function only for the first gap of each intersection and is ignored for all the rest. It is used to capture the behaviour of not stopping at all at the junction and accepting the first gap – the likelihood of which is expected to increase if the driver is travelling at a high speed.

Finally, the dummy variable of the last gap (which is 1 if there are no further approaching vehicles on sight) has a negative effect denoting a reduction in the probability of accepting a gap which is the last one. This confirms that drivers' gap-acceptance decisions are not short-sighted or focused on the current gap only, rather, the drivers further consider the next available gaps before deciding whether to accept the immediately available gap or not (anticipation effect). The variable sign is thus intuitive.

Model 2: Socio-demographics variables included

Model 2 included all of the Model 1 variables as well as the variables related to the sociodemographic characteristics of the drivers. Among the several sociodemographic variables tested, those with a statistically significant effect are Age>45 (which is 1 if the driver is older than 45 years, 0 otherwise) and Regular driver dummy (which is 1 if the driver typically drives every day, 0 otherwise). It may be noted that these variables are used in the dummy variable form, since it provides a better model fit with this coding, rather than having a continuous or ordinal form. The Age>45 dummy has a negative effect on gap-acceptance probability, indicating that older drivers are less likely to accept an available gap compared to younger. Moreover, all else being equal, participants that drive every day are more likely to accept a gap. It may be noted that the effects of gender, accident records and fine for speeding have also been tested but not found to have a statistically significant effect. The signs of the variables common with Model 1 were found to be the same but the magnitudes were different. Such changes in sensitivity are expected as the socio-demographic variables are adding further insights in the observed behaviour potentially leading to more representative sensitivity values.

The results of the gap-acceptance model(s) of this study support the existing literature findings. For instance, previous research (e.g. Matthews et al., 1999) used driving frequency as a measure of driving exposure and positively related it to crashes and speed violations. In the present case, participants driving on a daily basis – and thus with higher exposure - were more likely to accept a gap and therefore might be considered as more risk-takers. Similarly, in existing research elder drivers are found to have a less risk-taking propensity (e.g. Jonah, 1990; Krahé and Fenske, 2002; Rhodes and Pivik, 2011; Taubman-Ben-Ari and Yehel, 2012) which is in agreement with our findings.

Model 3: Time pressure considered

The third gap-acceptance model (Model 3) includes all the variables of Model 2 and also accounts for the time pressure conditions induced at the second gap-acceptance task. The time pressure parameter has a positive effect indicating that drivers were more likely to accept a gap if they are subjected to time pressure. Again, the signs of the variables common with Model 2 were found to be the same but the magnitudes were different.

Model 4: Physiological variables included

Finally, the model is enhanced by physiological variables related to heart rate and SCRs. The extraction and transformation/normalization of the physiological responses is described in section 4.2. Both variables have a significant a positive effect. This outcome, together with the effect of time pressure conditions, confirm that drivers' (gap-acceptance) behaviour is not only influenced by traffic conditions but also by external stressors (time pressure in this case) or acute stress levels. In the current case, drivers' stress is reflected through physiological responses during gap-acceptance choices, where a rise in the indicator values also implies an increase in the probability of crossing. However, the crossing behaviour, as examined in the present study can be also interpreted as an action that involves risk-taking propensity. Drivers' physiological responses can hence be seen as indicators of potential aberrant or risky behaviour that could lead to a crash.

The main findings of the presented models are in accordance with literature findings, in

terms of the effect of gap size on drivers' behaviour, as participants were more likely to accept larger gaps. The effect of waiting time was also investigated, but no statistically significant outcomes were found. Moreover, potential queuing effects were not examined as we controlled for this effect and there was no other traffic on the minor road. Finally, literature findings (e.g. Bottom & Ashworth, 1978; Nabaee et al., 2011) suggest that older drivers tend to accept larger gaps. This outcome is in line with our results since older drivers had a smaller probability of accepting a gap.

5.2. Model comparison

As shown in Table 6, while the gap-acceptance model is being enriched with new parameters, measures of model improve, both for the final log-likelihood (LL) and the ρ^2 and adjusted ρ^2 values.

All models are next compared using the likelihood ratio test (e.g. Ben-Akiva and Lerman, 1985). In brief, the test can be defined as:

$$LR = -2(LL^R - LL^U)$$

where L^R is the LL value of the restricted model (the one with fewer variables) and L^U is the LL of the unrestricted model (the model that includes the extra variables). The resulting LR statistic is asymptotically χ^2-distributed and is compared with a critical value which depends on the degrees of freedom (difference in estimated parameters). If the LR statistic exceeds that threshold value then the null hypothesis that both models perform equally is rejected.

The results of the various likelihood ratio tests are presented in Table 6. In all cases, the null hypothesis is rejected at 99% level which implies that the models with more variables have a significantly better goodness-of-fit compared to the simpler models re-confirming the hypotheses that driving is a complex task affected by factors beyond traffic conditions. Furthermore, since Model 4 has a significantly better goodness-of-fit compared to Model 3 - indicating statistically significant improvements in the model fit due to the incorporation of physiological variables.

Table 6: Likelihood ratio tests' results

Models	LR	Degrees of freedom (df)	$\chi^2_{(99\%,df)}$	Null hypothesis
Model 2 vs Model 1	15.41	2	9.21	Rejected
Model 3 vs Model 2	8.43	1	6.64	Rejected
Model 4 vs Model 3	13.42	2	9.21	Rejected

5.3 Sensitivity analysis

The effect of each variable on the gap-acceptance probabilities is investigated first. In this regard, each variable is varied within the predefined bounds (specified by the range of values observed in the experimental data) while keeping all other variables constrained to the sample averages. The fixed values of the continuous variables used are 4.295s for gap size, 0.96m/s for speed, 4.0543m for the position (median value), -0.15 for the normalised

heart rate and 0.038 for the normalised SCRs. For the dummy variables, sample average values are also used (varying between zero and one): 0.18 for age, 0.46 for driving frequency, 0.45 for time pressure and 0.05 for the last gap. Based on these values, the probabilities of gap-acceptance are estimated for the variables common in the Model 1[2] and the Model 4 (based on model fit results in section 5.2) as presented in Figure 6.

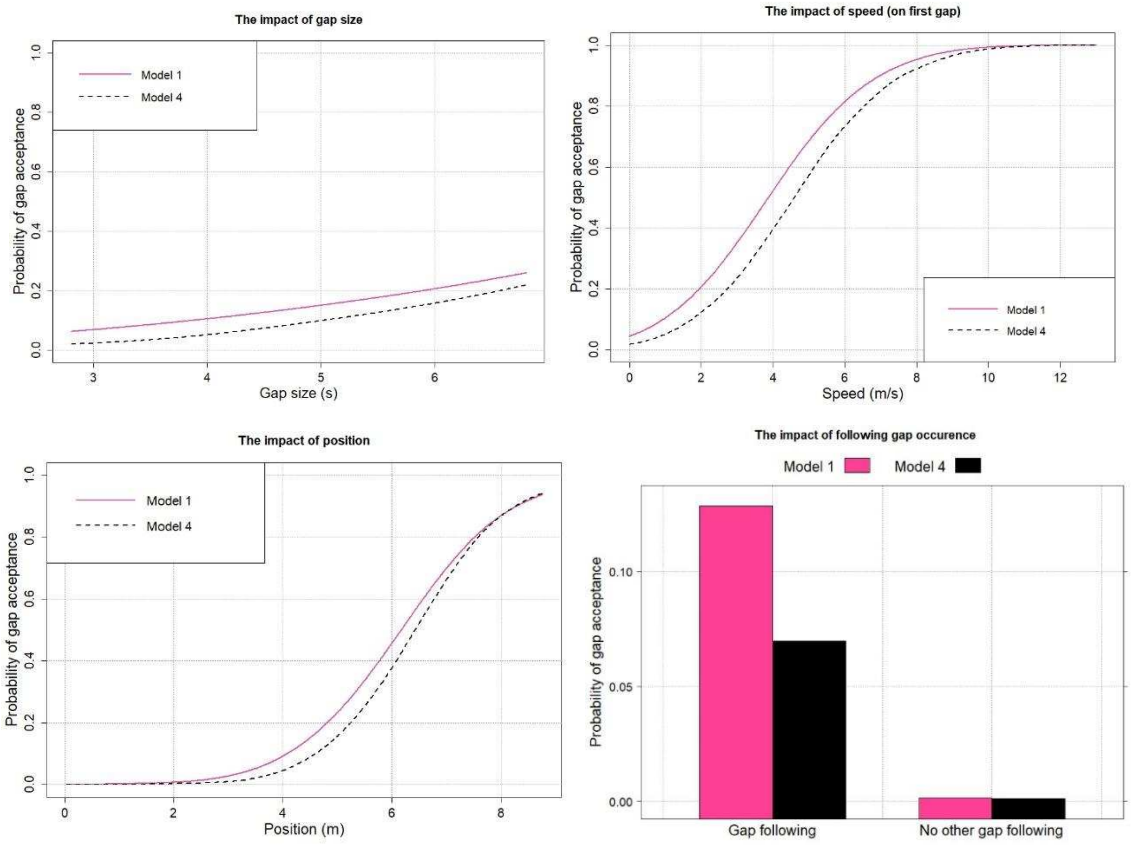

Figure 6: Variations in gap acceptance probabilities in Models 1 and 4

A general observation from Figures 6 is that in case of all traffic variables, the general trends of change in the probabilities are similar for both Model 1 and Model 4. For example, all else being equal, the probability of accepting a gap increases with gap size, speed (for the first gap), the position with respect to the intersection and the gap being the last gap. However, all else being equal, the probabilities of accepting a gap are always higher for Model 1, denoting overprediction of accepting a gap if the driver characteristics and stress levels are not included in the model.

Figure 7 depicts the effect of the socio-demographic variables used in the Model 4[3]. With respect to the age dummy variable, the probability for accepting a gap, for a driver above 45 years, has a value close to zero while gap-acceptance probability increases for younger

[2] It may be noted that the state-of-the-art traffic simulation tools are based on the principles of Model 1.

[3] Since these variables are not included in the Model 1, their effect on gap-acceptance probabilities have not been investigated across models but only for Model 4.

drivers. In a similar way, the gap-acceptance probabilities for participants who drive on a daily basis, are higher compared to the rest. Finally, as expected, the probability for accepting a gap under time pressure conditions is almost double compared to no time pressure.

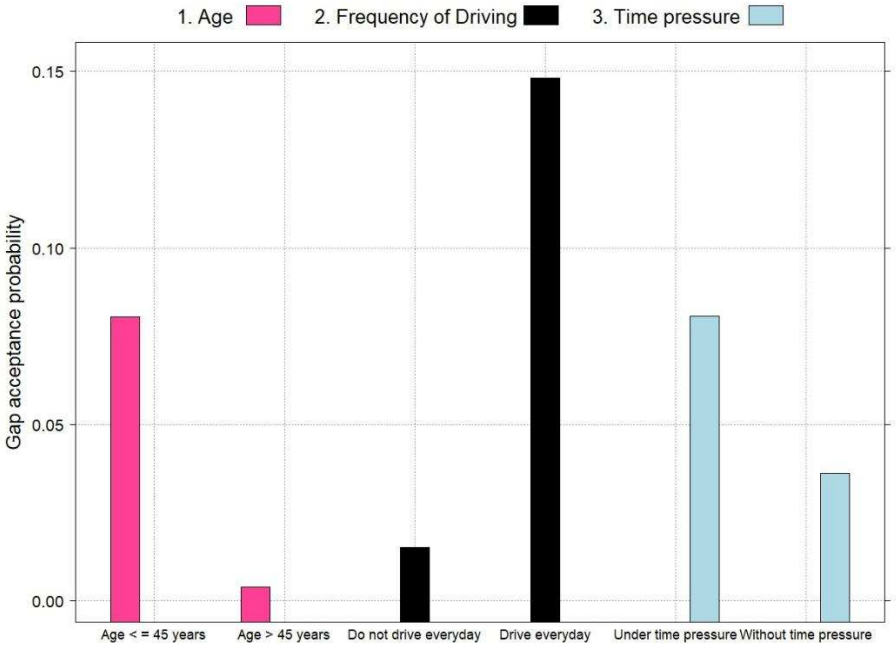

Figure 7: Sensitivity plots of the dummy variables used in Model 4 on gap-acceptance probability

The effect of the physiological measurement variables is shown in Figure 8. The results show that the gap acceptance probabilities increase in a similar pattern as the values of heart rate and increase in SCR.

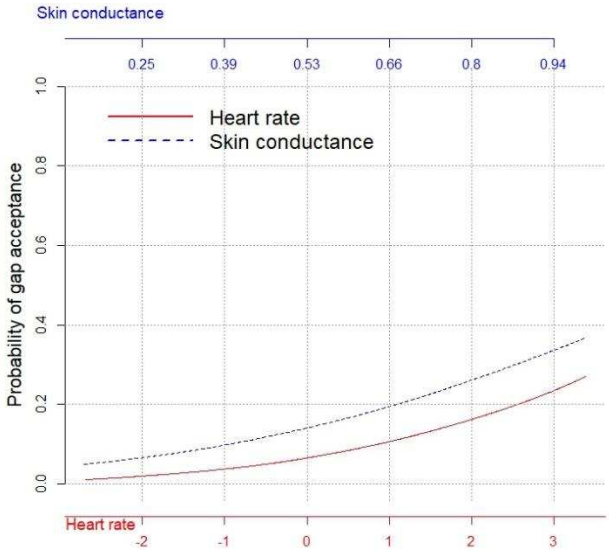

Figure 8: Sensitivity plot of heart rate and SCR on gap-acceptance probability

5.4 Substitution rates

At the final part of the analysis, an alternative approach is attempted to compare Model 1 and the Model 4. The approach is based on the marginal rates of substitutions (MRS) that also assists in avoiding issues of differences in scales across models. The MRS investigates the required change in a specific variable, in order to counterbalance the change in another variable and keep the total utility constant. The MRS is calculated as the ratio of the parameter estimates (β_i/β_j), where i and j denote two different variables of the model. In most studies, MRS has been used to calculate marginal willingness-to-pay, using the marginal utility of price in the denominator and another variable (travel time for instance) in the numerator. In this case, the parameter of gap size has been used as the denominator and the ratios are computed using each of the other parameters as numerators. The results are illustrated in Figure 9 where the calculated MRS values represent the relative effect of each parameter with respect to the gap size parameter in each model.

It should be mentioned that since the parameter of gap size is positive, the ratios with negative parameter are expected to be negative while positive ratios are expected when the opposite holds. Thus, when interpreting the MRS values, what is important is whether the absolute ratio value is higher than unity, rather than the sign of the ratio itself. For instance, |MRS|>1 shows that the change in utility, from a one-unit shift from the baseline of a given variable, is greater than the change corresponding to an increase in gap size by 1s. The opposite applies for |MRS|<1.

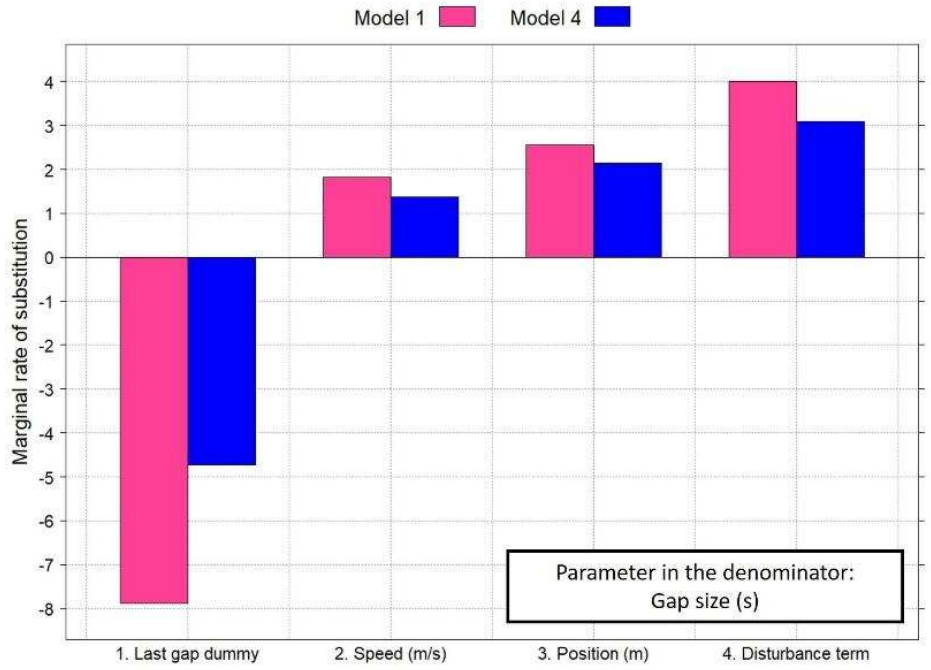

Figure 9: Marginal rates of substitution

As observed in Figure 9, the absolute values of MRS are larger than unity denoting all these variables have a higher contribution to the utility (in absolute terms) compared to the gap size variable (i.e. per second). Moreover, in all cases, the absolute values are higher for the

Model 1. The MRS for the last gap dummy indicates that the effect of the current gap being the last in the sequence of gaps is almost 8 times as negative as the increase of 1s in the gap size in the utility of gap acceptance in Model 1. However, in Model 4, it is 6 times as negative as the increase of 1s in gap size. For the approach speed, in the utility of gap acceptance in Model 1, the effect of an increase in approach speed of 1m/s is twice as positive as an increase of 1s in gap size. The same ratio in the Model 4 denotes that 1 m/s increase in approaching speed is approximately 1.4 times as positive as 1s gap increase. Likewise, the effect of a 1m increase in vehicle's position (denoting proximity to the start of the intersection) is approximately 3.5 times and 2 times as positive as a 1s increase in gap size for Model 1 and Model 4 respectively. For the individual specific error term, the MRS values indicate the contribution of these in the utility are 4 and 3 times more than the contribution of gap size in Model 1 and Model 4 respectively. This reduction is expected as Model 4 captures the heterogeneity among the driver by means of the socio-demographic and physiological sensor variables leading to a reduction in unobserved heterogeneity.

6. Conclusion

The results of both the statistical analyses and the discrete choice model indicate a significant impact of time pressure on the gap-acceptance decision. The time pressure variable has an expected positive sign denoting that also else being equal, the probability of accepting a gap more than doubles in presence of time pressure. As expected, increasing gap size has a positive effect in acceptance probability. Moreover, socio-demographics as age and driving frequency, influence gap-acceptance probability. The effects of gap size and age are in line with the findings of previous literature. Further, empirical analyses demonstrate that the explanatory power of the models increases when the models are augmented with EDA and heart rate data. The gap acceptance probability was found to increase non-linearly with the increase with the skin conductance response and heart rates resulting significant increase in the probability (up to 40%) of accepting a gap. In addition, using the choice modelling framework made it possible to quantify the impact of time pressure and stress on sensitivities towards the traffic-related variables. Results indicate that the inclusion of the physiological sensor measurements reduced the sensitivities towards the traffic-related variables, which can have important safety implications. These findings indicate the need for an additional dimension that should be considered in driving behaviour models for more realistic representation of reality.

Despite the promising nature of the results, there are some limitations in this study that can be investigated in future research. First of all, the data was collected in a simulated environment and thus may be behaviourally incongruent due the experimental nature. However, it is not possible to control the driving situation to isolate the stress effects in a field study. We are investigating the transferability of models developed using the driving simulator to the field in separate research (Papadimitrou and Choudhury 2017) and ways to correct for the potential scale differences (Paschalidis et al. 2018) which will help to make the model coefficients more applicable in the field. Secondly, it should be noted that time pressure was always induced at the second intersection without counterbalancing between the two tasks. Though this is a standard approach in stress research (see Rendon-Velez et al., 2016 for example) and the learning effect is likely to be minimal given the

experimental design, this is yet to be tested empirically. Moreover, it is worth mentioning that physiological responses actually represent 'arousal' which may be a reflection of other emotional states, positive or negative. Given the experimental setting of the current study, and the expected impact on drivers' behaviour, we decided to conceptualize physiological responses as an expression of stress though it can be confounded with other forms of arousals as well. Another potential source of bias could be self-selection however, it is very likely that it is uncorrelated with stress levels and thus does not affect the results. Finally, in terms of the model structure, there is scope to use more advanced model structures (e.g. treating stress as a latent variable for instance) as well as enhance the models with 'life stress' and 'trait stress' data. Development of other driving behaviour models (signal violation, overtaking) and cross-comparison of the stress effects across scenarios will also be an interesting direction for future research.

In terms of practical application of the models for prediction, the challenge lies in inferring the presence of time pressure and/or stress levels in real-life driving. However, with advances in ubiquitous computing technologies, it is now becoming feasible to measure stress levels in a very non-intrusive manner – wearable wristbands (as used in this study) and smartphone technologies that can detect stress levels from pitch and intervals of voice conversations (Sharma and Gedeon 2012, Lu et al. 2012). Given the extremely steep growth rate of wearables and smartphones, as well as advent of semi-autonomous cars (which have a wide range of sensors for inferring the surrounding traffic conditions), it is likely to be possible in near future to establish sophisticated models to sense stress levels of the driver and correlate it with potential influencing factors. Such prediction models for stress levels in real-world conditions will be very useful in widespread applications of the proposed model. This, coupled with the advances in the field of artificial emotional intelligence (Emotion AI) which has made it possible to device interventions to reduce stress (Fletcher et al. 2010, Picard et al. 2011), can make a significant contribution in increasing road safety. For instance, advances in vehicle operation technologies offer the opportunity for designing interventions to warn/advise drivers, limit acceleration-deceleration capabilities, introduce calming measures and even take over full control of the vehicle. The proper value addition of such novel technologies requires quantification of the safety impacts of stress. Our models can be used for such evaluations and/or subsequent willingness-to-pay. Applications may be also extended in the field of microsimulation to capture and better reflect driver heterogeneity. For example, there are emerging microsimulation models that combine activity models with traffic microsimulation (e.g. SimMobility (Adnan et al. 2016)). In these new types of tools, it is possible to include schedule delays in the traffic simulation component and our models can contribute to more realistic representation of driving behaviour in such simulation tools and hence increase their accuracy.

ACKNOWLEDGEMENT

The core component of this research is supported by the Next Generation Driving Behaviour Model (NG-DBM) project funded by FP7Marie Curie Career Integration Grant of the European Union (PCIG14-GA-2013-631782). Professor Stephane Hess' time was supported by the European Research Council through the consolidator grant 615596-DECISIONS. We would like to thank Michael Daly of UoLDS team for creating the

driving simulator scenarios and Dr Daryl Hibberd and Professor Samantha Jamson for their feedback on the design of the experiments.

REFERENCES

Abou-Zeid, M., & Ben-Akiva, M. (2014). 17 Hybrid choice models. *Handbook of choice modelling*, 383.

Ahmed, K. I. (1999). *Modeling drivers' acceleration and lane changing behavior* (Doctoral dissertation, Massachusetts Institute of Technology).

Ahmed, N., Iftekhar, L., Ahmed, S., Rahman, R., Reza, T., Shoilee, S., & Choudhury, C. F. (2015, December). Bap re Bap!: Driving Experiences through Multimodal Unruly Traffic on Bumpy Roads. In *Proceedings of the 2015 Annual Symposium on Computing for Development* (pp. 63-64). ACM.

Amin, H. J., & Maurya, A. K. (2015). A review of critical gap estimation approaches at uncontrolled intersection in case of heterogeneous traffic conditions. *Journal of transport literature*, *9*(3), 5-9.

Ashton, W. D. (1971). Gap-acceptance problems at a traffic intersection. *Applied Statistics*, 130-138.

Ashworth, R. (1969). The capacity of priority-type intersections with a non-uniform distribution of critical acceptance gaps. *Transportation Research*, *3*(2), 273-278.

Adnan, M., Pereira, F.C., Azevedo, C.M.L., Basak, K., Lovric, M., Raveau, S., Zhu, Y., Ferreira, J., Zegras, C. and Ben-Akiva, M., 2016. Simmobility: A multi-scale integrated agent-based simulation platform. In 95th Annual Meeting of the Transportation Research Board Forthcoming in Transportation Research Record.

Barceló, J. (2010). *Fundamentals of traffic simulation* (Vol. 145, p. 439). New York: Springer.

Ben-Akiva, M. E., & Lerman, S. R. (1985). *Discrete choice analysis: theory and application to travel demand* (Vol. 9). MIT press.

Benedek, M., & Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of neuroscience methods*, *190*(1), 80-91.

Ben‐Shakhar, G. (1985). Standardization within individuals: A simple method to neutralize individual differences in skin conductance. *Psychophysiology*, *22*(3), 292-299.

Bottom, C. G., & Ashworth, R. (1978). Factors affecting the variability of driver gap-acceptance behaviour. *Ergonomics*, *21*(9), 721-734.

Brilon, W., Koenig, R., & Troutbeck, R. J. (1999). Useful estimation procedures for critical gaps. *Transportation Research Part A: Policy and Practice*, *33*(3), 161-186.

Cacioppo, J. T., Tassinary, L. G., & Berntson, G. (Eds.). (2007). *Handbook of psychophysiology*. Cambridge University Press.

Choice Modelling Centre, University of Leeds. URL: https://cmc.leeds.ac.uk/

Choudhury, C. F. (2007). *Modeling driving decisions with latent plans* (Doctoral dissertation, Massachusetts Institute of Technology).

Clapp, J. D., Olsen, S. A., Danoff-Burg, S., Hagewood, J. H., Hickling, E. J., Hwang, V. S., & Beck, J. G. (2011). Factors contributing to anxious driving behavior: The role of stress history and accident severity. *Journal of anxiety disorders*, *25*(4), 592-598.

De Winter, J. C. (2013). Using the Student's t-test with extremely small sample sizes. *Practical Assessment, Research & Evaluation*, *18*(10).

Empatica, https://www.empatica.com/

Fitzpatrick, K. (1991). Gaps accepted at stop-controlled intersections. *Transportation Research Record*, *1303*, 103-112.

Farah, H., Bekhor, S., Polus, A. and Toledo, T., 2009. A passing gap acceptance model for two-lane rural highways. Transportmetrica, 5(3), pp.159-172.

Farah, H., & Koutsopoulos, H. N. (2014). Do cooperative systems make drivers' car-following behavior safer?. *Transportation research part C: emerging technologies*, *41*, 61-72.

Fletcher, R. R., Dobson, K., Goodwin, M. S., Eydgahi, H., Wilder-Smith, O., Fernholz, D., ... & Picard, R. W. (2010). iCalm: Wearable sensor and network architecture for wirelessly communicating and logging autonomic activity. *IEEE Transactions on Information Technology in Biomedicine*, *14*(2), 215-223.

Garrity, R. D., & Demick, J. (2001). Relations among personality traits, mood states, and driving behaviors. *Journal of Adult Development*, *8*(2), 109-118.

Ge, Y., Qu, W., Jiang, C., Du, F., Sun, X., & Zhang, K. (2014). The effect of stress and personality on dangerous driving behavior among Chinese drivers. *Accident Analysis & Prevention*, *73*, 34-40.

Greenshields, B. D., Schapiro, D., & Ericksen, E. L. (1946). *Traffic performance at urban street intersections* (No. Tech Rpt 1).

Gulian, E., Matthews, G., Glendon, A. I., Davies, D. R., & Debney, L. M. (1989). Dimensions of driver stress. *Ergonomics*, *32*(6), 585-602.

Guo, R. J., Wang, X. J., & Wang, W. X. (2014). Estimation of critical gap based on Raff's definition. *Computational intelligence and neuroscience*, *2014*, 16.

Halton, J. H. (1960). On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik*, *2*(1), 84-90.

Healey, J. A., & Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on intelligent transportation systems*, *6*(2), 156-166.

Hennessy, D. A., & Wiesenthal, D. L. (1999). Traffic congestion, driver stress, and driver aggression. *Aggressive behavior*, *25*(6), 409-423.

Hill, J. D., & Boyle, L. N. (2007). Driver stress as influenced by driving maneuvers and

roadway conditions. *Transportation Research Part F: Traffic Psychology and Behaviour*, *10*(3), 177-186.

Hoogendoorn, R., Hoogendoorn, S., Brookhuis, K., & Daamen, W. (2010). Mental workload, longitudinal driving behavior, and adequacy of car-following models for incidents in another driving lane. *Transportation Research Record: Journal of the Transportation Research Board*, (2188), 64-73.

Jamson, S., Lai, F., & Jamson, H. (2010). Driving simulators for robust comparisons: A case study evaluating road safety engineering treatments. *Accident Analysis & Prevention*, *42*(3), 961-971.

Jonah, B. A. (1990). Age differences in risky driving. *Health Education Research*, *5*(2), 139-149.

Kalimeri, K., & Saitis, C. (2016, October). Exploring multimodal biosignal features for stress detection during indoor mobility. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction* (pp. 53-60). ACM.

Karenbach, C. (2005). Ledalab-a software package for the analysis of phasic electrodermal activity. *Technical Report, Allgemeine Psychologie, Institut für Psychologie, Tech. Rep.*

Katsis, C. D., Katertsidis, N. S., & Fotiadis, D. I. (2011). An integrated system based on physiological signals for the assessment of affective states in patients with anxiety disorders. *Biomedical Signal Processing and Control*, *6*(3), 261-268.

Kontogiannis, T. (2006). Patterns of driver stress and coping strategies in a Greek sample and their relationship to aberrant behaviors and traffic accidents. *Accident Analysis & Prevention*, *38*(5), 913-924.

Krahé, B., & Fenske, I. (2002). Predicting aggressive driving behavior: The role of macho personality, age, and power of car. *Aggressive Behavior*, *28*(1), 21-29.

Lu, H., Frauendorfer, D., Rabbi, M., Mast, M. S., Chittaranjan, G. T., Campbell, A. T., ... & Choudhury, T. (2012, September). Stresssense: Detecting stress in unconstrained acoustic environments using smartphones. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (pp. 351-360). ACM.

Lykken, D. T. (1972). Range correction applied to heart rate and to GSR data. *Psychophysiology*, *9*(3), 373-379.

Maaoui, C., & Pruski, A. (2010). Emotion recognition through physiological signals for human-machine communication. In *Cutting Edge Robotics 2010*. InTech.

Mahmassani, H., & Sheffi, Y. (1981). Using gap sequences to estimate gap acceptance functions. *Transportation Research Part B: Methodological*, *15*(3), 143-148.

Marczak, F., Daamen, W., & Buisson, C. (2013). Key variables of merging behaviour: empirical comparison between two sites and assessment of gap acceptance theory. *Procedia-Social and Behavioral Sciences*, *80*, 678-697.

Markkula, G. M., Romano, R., Jamson, A. H., Pariota, L., Bean, A., & Boer, E. R. (2018). Using driver control models to understand and evaluate behavioural validity of driving

simulators. *IEEE Transactions on Human-Machine Systems*.

Matthews, G., Tsuda, A., Xin, G., & Ozeki, Y. (1999). Individual differences in driver stress vulnerability in a Japanese sample. *Ergonomics*, *42*(3), 401-415.

Maze, T. H. (1981). A probabilistic model of gap acceptance behavior. *Transportation research record*, *795*, 8-13.

Miller, A. J., and R. L., Pretty (1968), Overtaking on two-lane rural roads. *Proc. Aust. Rd. Res. Board*, Vol.4, No.1, 582–591.

Nabaee, S., Moore, D., & Hurwitz, D. (2011). Revisiting Driver Behavior at Unsignalized Intersections: Time of Day Implications for Two-Way Left Turn Lanes (TWLTL).

Ossen, S., & Hoogendoorn, S. (2005). Car-following behavior analysis from microscopic trajectory data. *Transportation Research Record: Journal of the Transportation Research Board*, (1934), 13-21.

Papadimitriou, S. and Choudhury, C.F. (2017). Transferability of Car-Following Models Between Driving Simulator and Field Traffic. Transportation Research Record: Journal of the Transportation Research Board, (2623), pp.60-72.

Paschalidis, E., Choudhury, C.F. and Hess, S. (2018). Improving the Transferabilty of Car-Following Models Between Driving Simulator and Field Traffic, 97th Annual Meeting of the Transportation Research Board, USA.

Picard, R. W., Vyzas, E., & Healey, J. (2001). Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE transactions on pattern analysis and machine intelligence*, *23*(10), 1175-1191

Pollatschek, M. A., Polus, A., & Livneh, M. (2002). A decision model for gap acceptance and capacity at intersections. *Transportation Research Part B: Methodological*, *36*(7), 649-663.

Qu, W., Ge, Y., Jiang, C., Du, F., & Zhang, K. (2014). The Dula Dangerous Driving Index in China: an investigation of reliability and validity. *Accident Analysis & Prevention*, *64*, 62-68.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Raff, M.S., J.W. Hart (1950), A volume warrant for urban stop signs. Eno Foundation for Highway Traffic Control, Saugatuck, Connecticut, USA.

Rendon-Velez, E., Van Leeuwen, P. M., Happee, R., Horváth, I., Van der Vegte, W. F., & De Winter, J. C. F. (2016). The effects of time pressure on driver performance and physiological activity: a driving simulator study. *Transportation research part F: traffic psychology and behaviour*, *41*, 150-169.

Rigas, G., Goletsis, Y., & Fotiadis, D. I. (2012). Real-time driver's stress event detection. *IEEE Transactions on intelligent transportation systems*, *13*(1), 221-234.

Rhodes, N., & Pivik, K. (2011). Age and gender differences in risky driving: The roles of positive affect and risk perception. *Accident Analysis & Prevention*, *43*(3), 923-931.

Sano, A., Picard, R. W., & Stickgold, R. (2014). Quantitative analysis of wrist electrodermal activity during sleep. *International Journal of Psychophysiology*, *94*(3), 382-389.

Sharma, N., & Gedeon, T. (2012). Objective measures, sensors and computational techniques for stress recognition and classification: A survey. *Computer methods and programs in biomedicine*, *108*(3), 1287-1301.

SimMobility–Integrated Simulation Platform URL: https://its.mit.edu/software/simmobility

Singh, M., & Queyam, A. B. (2013). A novel method of stress detection using physiological measurements of automobile drivers. *International Journal of Electronics Engineering*, *5*(2), 13-20.

Singh, R. R., Conjeti, S., & Banerjee, R. (2013). A comparative evaluation of neural network classifiers for stress level analysis of automotive drivers using physiological signals. *Biomedical Signal Processing and Control*, *8*(6), 740-754.

Taubman-Ben-Ari, O., & Yehiel, D. (2012). Driving styles and their associations with personality and motivation. *Accident Analysis & Prevention*, *45*, 416-422.

Thiffault, P., & Bergeron, J. (2003). Monotony of road environment and driver fatigue: a simulator study. *Accident Analysis & Prevention*, *35*(3), 381-391.

Toledo, T. (2003). *Integrating driving behavior modeling* (Doctoral dissertation, Massachusetts Institute of Technology).

Toledo, T. (2007). Driving behaviour: models and challenges. *Transport Reviews*, *27*(1), 65-84.

Tupper, S. M., Knodler Jr, M. A., & Hurwitz, D. S. (2011). Connecting gap acceptance behavior with crash experience. In *3rd International Conference on Road Safety and SimulationPurdue UniversityTransportation Research Board*.

University of Leeds, http://www.leeds.ac.uk/

University of Leeds Driving Simulator (UoLDS), http://www.uolds.leeds.ac.uk/

Useche, S., Serge, A., & Alonso, F. (2015). Risky Behaviors and Stress Indicators between Novice and Experienced Drivers. *American Journal of Applied Psychology*, *3*(1), 11-14.

Westerman, S. J., & Haigney, D. (2000). Individual differences in driver stress, error and violation. *Personality and Individual Differences*, *29*(5), 981-998.

World Health Organization (2015). Violence, Injury Prevention and World Health Organization, Global Status Report on Road Safety.

Young, K., Regan, M., & Hammer, M. (2007). Driver distraction: A review of the literature. Distracted driving, 379-405.

Zhai, J., & Barreto, A. (2006, May). Stress Recognition Using Non-invasive Technology. In FLAIRS Conference (pp. 395-401).