Investigating the effect of driver-vehicle-environment interaction with risk through naturalistic driving data

Eva Michelaraki[a1,] Thodoris Garefalakis[a], Stella Roussou[a], Amir Pooyan Afghari[b], Evita Papazikou[c], Rachel Talbot[c], Muhammad Adnan[d], Muhammad Wisal Khattak[d], Christelle Al Haddad[e], Md Rakibul Alam[e], Constantinos Antoniou[e], Eleonora Papadimitriou[b], Tom Brijs[d], George Yannis[a]

[a]National Technical University of Athens, Department of Transportation Planning and Engineering, 5 Heroon Polytechniou str., Athens, GR-15773, Greece
[b]Delft University of Technology, Safety and Security Science Group, Jaffalaan 5, 2628 BX Delft, The Netherlands
[c]School of Design & Creative Arts, Transport Safety Research Centre, Loughborough University, Loughborough, Leicestershire LE11 3TU, UK
[d]UHasselt, School for Transportation Sciences, Transportation Research Institute (IMOB), Agoralaan, 3590 - Diepenbeek, Belgium
[e]Technical University of Munich, Chair of Transportation Systems Engineering, Arcisstrasse 21, 80333, Munich, Germany

Abstract

While mobility and safety of drivers are challenged by behavioural changes, the increasingly complex road environment has placed a higher demand on their adaptability. The ultimate goal of this paper was to identify the impact that the balance between task complexity and coping capacity had on crash risk. Towards that aim, an integrated model for understanding the effect of the inter-relationship of task complexity and coping capacity with risk was developed. A vast library of data from a naturalistic driving experiment was created in three countries (i.e. Belgium, UK and Germany) to investigate the most prominent driving behaviour indicators available, including speeding, headway, overtaking, duration, distance and harsh events. In order to fulfil the aforementioned objectives, exploratory analysis, such as Generalized Linear Models (GLMs) were developed and the most appropriate variables associated to the latent variable "task complexity" and "coping capacity" were estimated from the various indicators. Additionally, Structural Equation Models (SEMs) were used to explore how the model variables were inter-related, allowing for both direct and indirect relationships to be modelled. The analyses revealed that higher task complexity levels lead to higher coping capacity by drivers. Additionally, the effect of task complexity on risk was greater than the impact of coping capacity in Belgium and Germany, while mixed results were observed in the UK.

Keywords: driving behaviour; road safety; naturalistic driving experiment; Structural Equation Models; Generalized Linear Models.

* Corresponding author. Tel.: +030-210-772-1265;
E-mail address: evamich@mail.ntua.gr

1. Introduction

Ensuring road safety is paramount, aiming to reduce crash risk, prevent injuries, and save lives. Every year, a significant number of lives are lost, and many people suffer severe injuries due to road crashes. Multiple factors exert a substantial influence on road safety, potentially leading to crashes and affecting the seriousness of resulting injuries. Human behaviour, for example, assumes a pivotal role in road safety. Elements such as speeding, distracted driving (e.g., mobile phone use), impaired driving (i.e. caused by alcohol, drugs, or fatigue), aggressive driving, and failure to adhere to traffic regulations can elevate crash risk. Moreover, the design, state, and upkeep of roadways and infrastructure also play a role in road safety. Inadequate road design, insufficient signage, the absence of pedestrian crossings, insufficient lighting, and subpar maintenance can all contribute to crashes and injuries.

Simultaneously, the state and safety features of vehicles exert a substantial influence on road safety. Aspects like vehicle upkeep, tire condition, brake performance, and the presence of safety technologies can have a significant impact on the outcomes of crashes. Similarly, environmental conditions can have repercussions on road safety. Elements such as adverse weather conditions (e.g., rain, snow, fog), diminished visibility, and uneven road surfaces can elevate the likelihood of crashes. Additionally, socioeconomic factors, including income level, education, and access to transportation resources, can indirectly shape road safety. Disparities in these factors may give rise to variations in driver behaviours, vehicle conditions, and the quality of road infrastructure.

Based on the above, the overall goal of the i-DREAMS project is to establish a framework for defining, developing, testing, and validating a context-aware safety framework for driving, referred to as the "Safety Tolerance Zone." This framework is integrated within a smart Driver, Vehicle & Environment Assessment and Monitoring System (i-DREAMS). By considering various factors related to the driver's background, real-time risk indicators linked to driving performance, driver condition, and the complexity of the driving task, a continuous, real-time assessment is conducted to determine if a driver is operating within safe parameters, known as the "Safety Tolerance Zone".

According to the level of unsafe driving behaviour, the STZ is categorized into three levels: ‘Normal’, ‘Dangerous’ and ‘Avoidable Accident’. Firstly, the ‘Normal’ level denotes a situation with a minimal crash risk and thus safe driving practices. Secondly, the 'Dangerous' level refers to the chance of a crash increasing, but the crash is not unavoidable. Finally, the ‘Avoidable Accident’ level denotes a high risk of a potential crash occurring, but there is still enough time for drivers to act and avoid the incident.

Following the i-DREAMS project’s goal, this study aims to investigate the interaction between task complexity and coping capacity (i.e., related to both vehicle state and operator state factors). To achieve this goal, a complete Structural Equation Model (SEM) developed and a set of quantitative effects of indicators was created, describing the impacts of vehicle, operator and context characteristics on risk under different conditions. Apart from SEMs, Generalized Linear Models (GLMs) were also used and the goodness-of-fit-metrics for the models were explained.

The paper is structured as follows. At the beginning, a detailed overview of the project and its overall objective is provided. Following that, a comprehensive literature review on the statistical analysis of driving behaviour is presented. Furthermore, the data collection process is thoroughly described. The research approach is then outlined, including the theoretical foundations of the models used. Lastly, the results are provided, followed by substantial conclusions about the relationship between crucial factors such as task complexity and coping capacity on risk.

2. Background

The inter-relationship among task complexity, coping capacity, and crash risk is a multifaceted and crucial area of study in traffic safety research. The assessment of task difficulty and coping ability forms the basis of the i-DREAMS platform.

To begin with, task complexity plays a significant role in influencing crash risk on the roads. The complexity of driving tasks refers to the level of cognitive demand and physical effort required to perform them. Factors contributing to task complexity include traffic density, road infrastructure, weather conditions, presence of distractions, and time pressure, among others. The current state of the real-world environment in which a vehicle is being driven is related to task complexity. The registration of road layout (i.e., highway, rural, urban), time and place, traffic volumes (i.e., high, medium, low), and weather is particularly used to assess job complexity context.

On the other hand, coping capacity refers to an individual driver's ability to effectively manage and adapt to complex driving tasks. It encompasses factors such as experience, skills, perceptual abilities, decision-making processes, and the availability of appropriate coping strategies. Drivers with high coping capacity can better handle complex tasks, maintain situational awareness, and make appropriate decisions to mitigate crash risk. The conceptual foundation for the prediction of risk as a function of coping capacity and task complexity is shown in Figure 1.

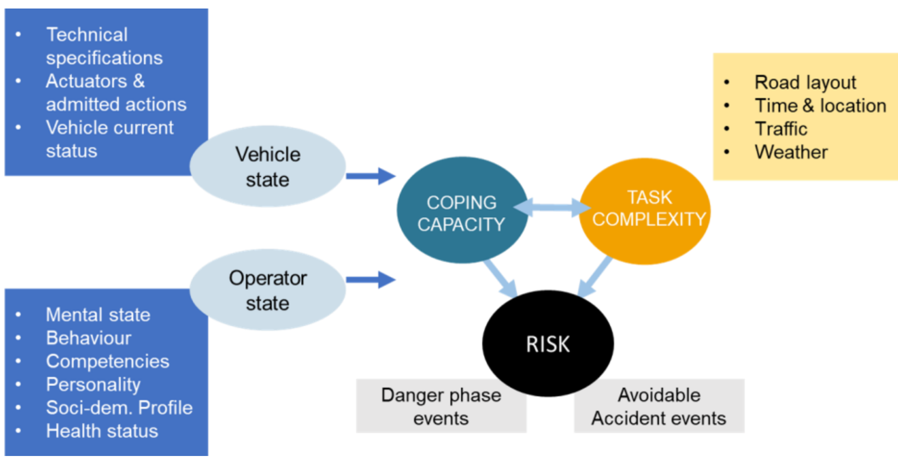

Figure 1: Post-hoc prediction of risk in function of coping capacity and task complexity

Road safety is a pressing global concern, with millions of lives lost or impacted by traffic crashes each year. To effectively address this issue, researchers and policymakers have turned to advanced statistical modelling techniques to gain a deeper understanding of the complex relationships between various factors contributing to road crashes.

In particular, SEMs have emerged as a powerful tool for analysing the intricate interplay between observed variables and latent constructs in road safety research. They allow researchers to explore the direct and indirect effects of multiple factors on road safety while providing a methodology for direct modelling of latent variable, separating measurement errors from true scores of attributes (Yuan & Bentler, 2006). This makes SEMs particularly suitable for studying the multifaceted nature of road safety, where numerous factors interact to influence the occurrence and severity of crashes. The application of SEMs in recent road safety research has yielded valuable insights into the underlying factors contributing to crashes and their consequences. By modelling and examining the relationships between various risk factors, SEMs help researchers identify key predictors of road crashes, understand their interrelationships, and develop effective intervention strategies (Shah et al., 2018).

Thus, the use of SEMs has proven invaluable in advancing road safety research. These models provide a comprehensive framework for understanding the intricate relationships and interdependencies among various factors contributing to road crashes. By elucidating causal mechanisms and mediating/moderating effects, SEMs enable researchers to develop targeted interventions, evaluate policy effectiveness, and ultimately enhance road safety outcomes.

3. Data Description

A naturalistic driving experiment was carried out involving 133 drivers from Belgium, UK and Germany and a large database of 26,908 trips and 500,000 minutes was created to investigate the most prominent driving behaviour indicators, including speeding, headway, duration, distance, and harsh acceleration and harsh brakings. The total number of drivers, trips and minutes is presented in Figure 2.

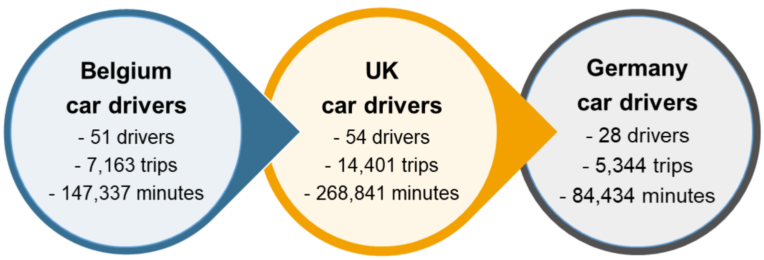

Figure 2: Number of drivers, trips, and minutes per country

Four separate SEM models were estimated in order to explore the relationship between the latent variables of task complexity, coping capacity and risk (expressed as the three stages of the STZ) of speeding and headway (level 1 'normal driving' used as the reference case). Figure 3 provides an overview of the different phases of the experimental design of the i-DREAMS on-road study.

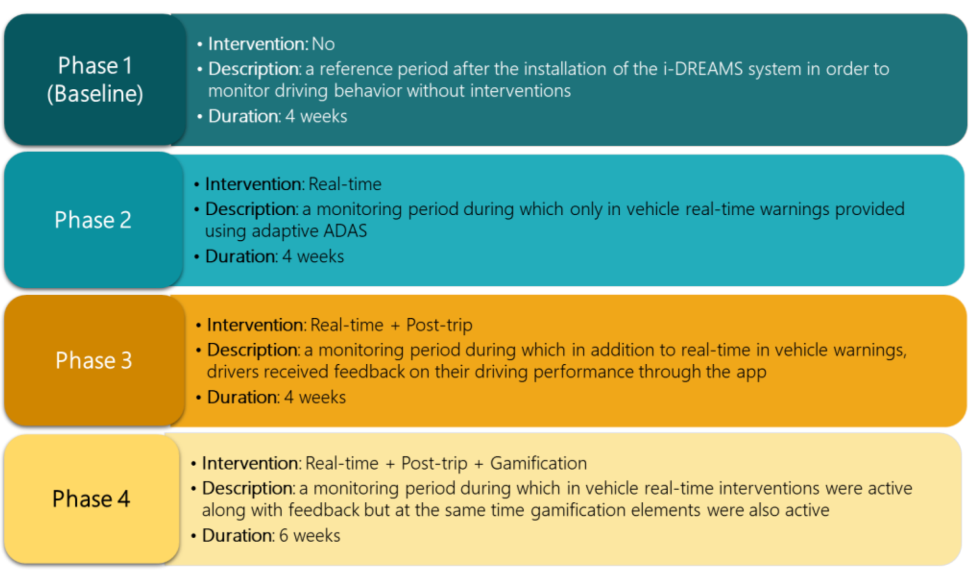

Figure 3: Overview of the different phases of the experimental design

4. Methodology

In order to fulfil the objectives of this study, exploratory analysis, such as Generalized Linear Models (GLMs) were developed and the most appropriate variables associated to the latent variable "task complexity" and "coping capacity" were estimated from the various indicators. In addition, SEMs were used to explore how the model variables were inter-related, allowing for both direct and indirect relationships to be modelled.

4.1 Generalized Linear Models (GLMs)

In statistics, the GLM is a flexible generalization of ordinary linear regression that allows for response variables that have error distribution models other than a normal distribution. The GLM generalizes linear regression by allowing the linear model to be related to the response variable via a link function and by allowing the magnitude of the variance of each measurement to be a function of its predicted value (Hastie & Pregibon, 2017).

In a GLM, each outcome Y of the dependent variables is assumed to be generated from a particular distribution in an exponential family, a large class of probability distributions that includes the normal, binomial, Poisson and gamma distributions, among others. The mean, μ, of the distribution depends on the independent variables, X, through:

$$E(Y|\mathrm{X}) = \mu = g^{-1}(X\beta) \quad \textbf{(1)}$$

where: E(Y|X) is the expected value of Y conditional on X; Xβ is the linear predictor, a linear combination of unknown parameters β; g is the link function.

In this framework, the variance is typically a function, V, of the mean:

$$Var(Y|\mathrm{X}) = V(g^{-1}(X\beta)) \quad \textbf{(2)}$$

It is convenient if V follows from an exponential family of distributions, but it may simply be that the variance is a function of the predicted value.

The unknown parameters, β, are typically estimated with maximum likelihood, maximum quasi-likelihood, or Bayesian techniques.

GLMs were formulated as a way of unifying various other statistical models, including linear regression, logistic regression, and Poisson regression. In particular, Hastie & Tibshirani (1990) proposed an iteratively reweighted least squares method for maximum likelihood estimation of the model parameters. Maximum-likelihood estimation remains popular and is the default method on many statistical computing packages. Other approaches, including Bayesian approaches and least squares fits to variance stabilized responses, have been developed.

A key point in the development of GLM was the generalization of the normal distribution (on which the linear regression model relies) to the exponential family of distributions. This idea was developed by Collins et al. (2001). Consider a single random variable y whose probability (mass) function (if it is discrete) or probability density function (if it is continuous) depends on a single parameter θ. The distribution belongs to the exponential family if it can be written as follows:

$$f(y;\theta) = s(y)t(\theta)e^{a(y)b(\theta)} \quad \textbf{(3)}$$

where: a, b, s, and t are known functions. The symmetry between y and θ becomes more evident if the equation above is rewritten as follows:

$$f(y;\theta) = \exp\left[\alpha(y)b(\theta) + c(\theta) + d(y)\right] \quad \textbf{(4)}$$

where: $s(y)=\exp[d(y)]$ and $t(\theta)=\exp[c(\theta)]$

It should be mentioned that the Variance Inflation Factor (VIF) is a measure of the amount of multicollinearity in regression analysis. Multicollinearity exists when there is a correlation between multiple independent variables in a multiple regression model. The default VIF cut-off value is 5; only variables with a VIF less than 5 will be included in the model (VIF<5). However, in certain cases, even if VIF is less than 10, then it can be accepted.

4.2 Structural Equation Models (SEM)

Structural Equation Modelling (SEM) or path analysis is a multivariate method used to test hypotheses regarding the influences among interacting observed and unobserved variables (Harrison et al., 2007). The observed variables are measurable, while unobserved variables are latent constructs.

SEM consist of two components: a measurement model and a structural model. The measurement model is used to assess how well various observable exogenous variables can measure the latent variables, as well as the measurement errors associated with them. The structural model is used to investigate the relationships among the model variables, enabling the modeling of both direct and indirect linkages. In this regard, SEMs distinguish themselves from regular regression techniques by deviating from direct relationships between variables.

The general formulation of SEM is as follows (Washington et al., 2020):

$$\eta = \beta\eta + \gamma\xi + \varepsilon \quad \textbf{(5)}$$

where: η represents a vector of endogenous variables, ξ represents a vector of exogenous variables, β and γ are vectors of coefficients to be estimated, and ε represents a vector of regression errors.

The measurement models can be described as follows (Chen, 2007):

$$x = \Lambda x\xi + \delta, \text{ for the exogenous variables} \quad \textbf{(6)}$$
$$y = \Lambda y\eta + \zeta, \text{ for the endogenous variables} \quad \textbf{(7)}$$

where: x and δ represent vectors associated with the observed exogenous variables and their errors, while y and ζ are vectors represent vectors associated with the observed endogenous variables and their errors. Λx, Λy are structural coefficient matrices that capture the effects of the latent exogenous and endogenous variables on the observed variables.

4.3 Model goodness-of-fit measures

In the context of model selection, model Goodness-of-Fit measures consist an important part of any statistical model assessment. Several goodness-of-fit metrics are commonly used, including the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), the goodness-of-fit index (GFI), the (standardized) Root Mean Square Error Approximation (RMSEA), the Comparative Fit Index (CFI) and the Tucker-Lewis Index (TLI). Such criteria are based on differences between the observed and modelled variance-covariance matrices. The results of the models were evaluated by satisfying the following statistical tests: p-value<0.001, CFI > 0.90, TLI > 0.90 and RMSEA<0.05.

5. Results

5.1 GLM results

GLMs were employed to investigate the relationship of key performance indicator of speeding for Belgian, UK and German car drivers. The relationship between speeding and risk is widely recognized in the road safety community and as such, speeding is a commonly used dependent variable in transportation human factors research.

The first GLM investigated the relationship between the speeding and several explanatory variables of task complexity and coping capacity (operator state) in Belgium. In particular, the dependent variable of the developed model is the dummy variable "speeding", which is coded with 1 if there is a speeding event and with 0 if not. The model parameter estimates are summarized in Table 2.

Table 2: Parameter estimates and multicollinearity diagnostics of the GLM for Belgium

Variables	Estimate	Standard Error	z-value	Pr(|z|)	VIF
(Intercept)	3.668	0.043	85.768	< .001	-
Time indicator	0.908	0.078	11.683	< .001	1.882
Weather	0.009	4.217×10^{-4}	20.952	< .001	1.228
High beam - Off	-0.018	7.062×10^{-4}	-25.286	< .001	1.470
Harsh acceleration	2.661	0.181	14.689	< .001	1.013
Distance	-6.128×10^{-4}	7.273×10^{-5}	-8.426	< .001	1.678
Summary statistics					
AIC	17404.428				
BIC	17413.817				
Degrees of freedom	88377				

Based on Table 2, it can be observed that all explanatory variables are statistically significant at a 95% confidence level; there is no issue of multicollinearity as the VIF values are much lower than 5. With regard to the coefficients, it was revealed that the indicators of task complexity, such as time indicator and wipers were positively correlated with speeding. The former refers to the time of the day (day coded as 1, dusk coded as 2, night coded as 3) which means that higher speeding events occur at night compared to during the day. This may be due to fewer cars on the road, lower visibility, and a false sense of security that comes with driving in the dark. Interestingly, wipers (wipers off coded as 0, wipers on coded as 1) were also found to have a positive correlation with speeding which means that there are more speeding events during adverse (e.g. rainy) weather conditions. This may be due to the fact that wet and slippery roads can make it more difficult to maintain control of the vehicle.

Additionally, rain can reduce visibility and make it harder to see other cars or obstacles on the road. Taking into account the indicator of high beam (indicating lighting conditions; no high beam detected), a negative correlation was identified which means that when high beam was off - and, therefore, it was daytime - there were less speeding events. This finding comes in agreement with the previous argument with the indicator of time of the day that higher speeding events occur at night compared to the rest of the day.

Regarding the indicators of coping capacity - operator state, harsh accelerations had a positive relationship with the dependent variable (i.e. speeding), indicating that as the number of harsh acceleration increases, speeding also increases. This is a noteworthy finding of the current research as it confirms that harsh driving behaviour events present a statistically significant positive correlation with speeding. Lastly, total distance travelled was negatively correlated with speeding which may be due to the fact that the longer a person drives, the more fatigued they may become, causing them to drive slower and more cautiously.

The second GLM investigated the relationship between the speeding and several explanatory variables of task complexity and coping capacity in UK. The model parameter estimates are summarized in Table 3.

Table 3: Parameter estimates and multicollinearity diagnostics of the GLM for UK

Variables	Estimate	Standard Error	z-value	Pr(|z|)	VIF
(Intercept)	-3.824	0.014	-274.620	< .001	-
Duration	4.672×10^{-5}	7.877×10^{-7}	59.317	< .001	1.058
Harsh acceleration	-0.187	0.012	-15.377	< .001	1.014
Weather	-0.273	0.023	-11.713	< .001	1.008
High beam	0.128	0.078	1.635	0.102	1.002
Forward collision warning	10.603	2.479	4.276	< .001	1.001
Right lane departure warning	0.357	0.014	25.348	< .001	1.026
Distance	0.002	1.876×10^{-5}	117.628	< .001	1.072
Gender - Male	0.373	0.012	31.757	< .001	1.056
Summary statistics					
AIC	263599.548				
BIC	263610.743				
Degrees of freedom	537681				

It can be observed that all explanatory variables are statistically significant at a 95% confidence level (VIF is lower than 5). With regard to the coefficients, it was revealed that the indicators of coping capacity are all positively correlated with speeding except for harsh acceleration events that appear to be fewer when speeding occurs. The opposite happens with Forward Collision Warning (FCW) and Lane Departure Warning (LDW) events that appear to be higher in case of speeding. An increase in the trip duration and the distance travelled is associated with an increase in speeding events, as well. The use of wipers though is, as expected, negatively associated with speeding events. Gender was a significant variable in this model showing that male drivers (males coded as 0, females as 1), are possibly prone to speeding while the use of high beams also was connected with higher speeding events possibly due to lighter night hours traffic.

The third GLM investigated the relationship between the speeding and several explanatory variables of task complexity and coping capacity (vehicle and operator state) in Germany. The model parameter estimates are summarized in Table 4.

Table 4: Parameter estimates and multicollinearity diagnostics of the GLM for Germany

Variables	Estimate	Standard Error	z-value	Pr(|z|)	VIF
(Intercept)	1.105	0.057	19.549	< .001	-
Duration	0.003	3.414×10^{-5}	73.366	< .001	1.262
Distance	5.735×10^{-4}	3.723×10^{-5}	15.404	< .001	1.029
Harsh acceleration	1.282×10^{-4}	1.974×10^{-6}	64.951	< .001	1.222
Fuel type - Petrol	0.219	0.010	21.446	< .001	1.328
Vehicle Age	3.162×10^{-5}	3.340×10^{-6}	9.469	< .001	1.277
Gender - Female	-0.275	0.021	-13.025	< .001	1.256
Age	-0.003	0.001	-2.289	0.022	1.076
Drowsiness	1.009×10^{-5}	2.656×10^{-6}	3.800	< .001	1.113
Time indicator	8.547×10^{-5}	1.925×10^{-6}	44.405	< .001	1.080
High beam - On	0.817	0.059	13.963	< .001	1.073
Summary statistics					
AIC	127971.813				
BIC	127981.881				
Degrees of freedom	174299				

Based on Table 4, it can be observed that all explanatory variables are statistically significant at a 95% confidence level; there is no issue of multicollinearity (VIF is lower than 5). It was revealed that the indicators of task complexity, such as time and high beam (indicating lighting conditions; no high beam detected) were positively correlated with speeding. Regarding the indicators of coping capacity – vehicle state such as fuel type and vehicle age were positively correlated with speeding. Furthermore, it was demonstrated that indicators of coping capacity – operator state, such as harsh accelerations, distance, duration and drowsiness had a positive relationship with the dependent variable (i.e. speeding), indicating that as the values of the aforementioned independent variables increases, speeding also increases.

Taking into consideration socio-demographic characteristics, gender and age were negatively correlated with speeding. Results revealed that the vast majority of male drivers displayed less cautious behaviour during their trips and exceeded more often the speed limits than female drivers. It is also remarkable that the negative value of the “Age” coefficient implied that as the value of the variable increased (higher value indicates increased age and, therefore, increased years of participant’s experience), the speeding percentage was lower. Young drivers appeared to have a riskier driving behaviour than older drivers and were more prone to exceed the speed limits.

5.2 SEM results

In order to investigate the relationship between the latent variables of task complexity, coping capacity, and risk (represented as the three stages of the STZ), four distinct SEM models were developed.

5.2.1. Belgian cars

The latent variable risk is measured by means of the STZ levels for speeding (level 1 ‘normal driving’ used as the reference case), with positive correlations of risk with the STZ. The structural model between the latent variables shows some interesting findings: first, task complexity and coping capacity are inter-related with a positive correlation – albeit the magnitude of this correlation is very small. This positive correlation indicates that higher task complexity is associated with higher coping capacity implying that drivers' coping capacity increases as the complexity of driving task increases. The more complex the situation becomes as a result of speeding, the better the driver's coping capacity will become, for example because of increased alertness.

Coping capacity is associated with higher risk, which is an interesting finding. It could be assumed that higher coping capacity might reduce risk; however, the coping capacity indicators in our sample include static demographic and self-reported behaviour indicators and therefore are more representative of driver personality and general driving styles, and less so of the real-time operator state during the experiment. For instance, indicators related to the level of sleepiness, fatigue or distraction were either not available or not significant in this model. Therefore, it can be concluded that younger, more confident and less compliant drivers exhibited lower risk in this experiment, in terms of exceeding the STZ speeding boundaries – a finding which can be attributed to higher alertness and exposure in complex environments, without however taking into account the variations of their state during these trips. Figure 4 illustrates the results for each phase.

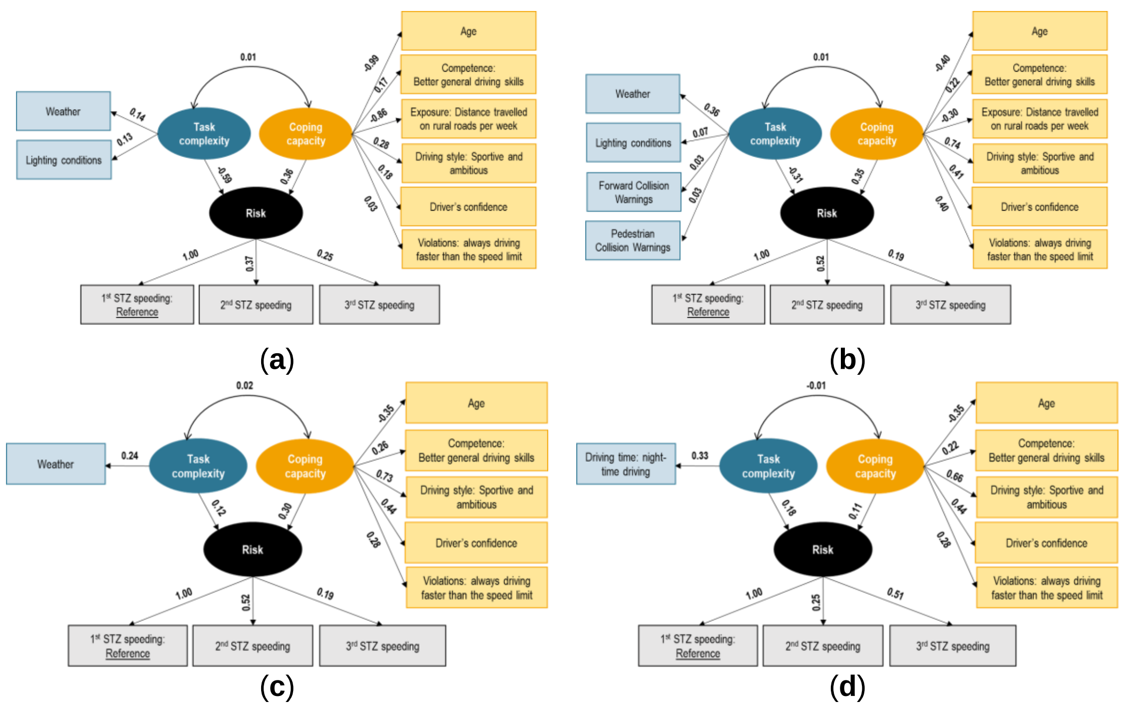

Figure 4: Results of SEM on risk – Belgian car drivers – experiment phase 1 (a), 2 (b), 3 (c), 4 (d).

The relationships between risk, task complexity, and coping capacity remain consistent across phases, with some noteworthy findings. In phase 2, FCW and PCW indicators load onto task complexity, reflecting real-time events that express demanding and risky situations. However, the overall impact of task complexity on risk only slightly decreases. These events may not be directly linked to exceeding speed limits, which defines risk in this context. Notably, these indicators aren't significant in the 3rd and 4th phases, likely due to lower event occurrences during these phases.

5.2.2. UK cars

Risk is measured by means of the STZ levels for headway (level 1 ‘normal driving’ used as the reference case; level 2 refers to ‘dangerous driving’, while level 3 refers to ‘avoidable accident driving’. In particular, negative correlations of risk with the STZ indicators were found.

The latent construct of task complexity is represented by the indicator variables of high beam and wipers use. Wipers can be an indication of weather conditions, most specifically, they can be indicative of rain presence during the trip while high beams can indicate lighting conditions, for example, low visibility or dark. Both variables have a positive loading on the latent factor task complexity showing that an increase in the latter explains an increase in both of them accordingly.

Regarding coping capacity, all the indicator variables in the model show a negative relationship with risk except for general sleeping rate. Driver style appears to be the most important indicator (higher estimate) for coping capacity and risk development while also important indicators are the speeding (driving always above speed limit), the mobile phone usage while driving, the illegal overtaking and the general sleeping rate. The latter, as expected, has a positive relationship with coping capacity showing that better sleep habits are associated with increased levels of driver capability. Last but not least, according to the model increased level of risks are linked to increased time spent on second and third headway level of STZ.

All the observed indicators presented in the model to represent the three latent concepts of task complexity, coping capacity and risk are statistically significant at 99.9% confidence level. Task complexity and coping capacity have a statistically significant impact on risk that is significantly interpreted by the time spent in each of the three levels of STZ regarding the headway indicator. As mentioned before in previous phases, lower risk relates to more time in the first level of STZ, in other words, to higher headways measurements. Similarly, to phase 4, task complexity has a greater effect (standardised coefficient=-0.26) on risk than coping capacity (standardised coefficient=-0.19).

In terms of the relationship between driving task complexity and risk the picture is different than in the other three phases. The model for phase 4 indicates that increased levels of driving task difficulty, related to weather and visibility conditions, are linked to lower levels of risk. This result could be interpreted by the fact that when drivers have to face more complicated road conditions such as rain or lower visibility, they could become more alerted and cautious.

Regarding the specific indicators of the latent concept of coping capacity, the same pattern can be observed as in all other phases with the driver style to dominate in the coping capacity latent construct. Furthermore, mobile phone use while driving, driving faster than the speed limit, driver style and illegal overtaking are all negatively related to coping capacity as it was intuitive. On the other hand, good sleeping rate is positively associated with driver capacity). The results for all phases are shown in Figure 5 below.

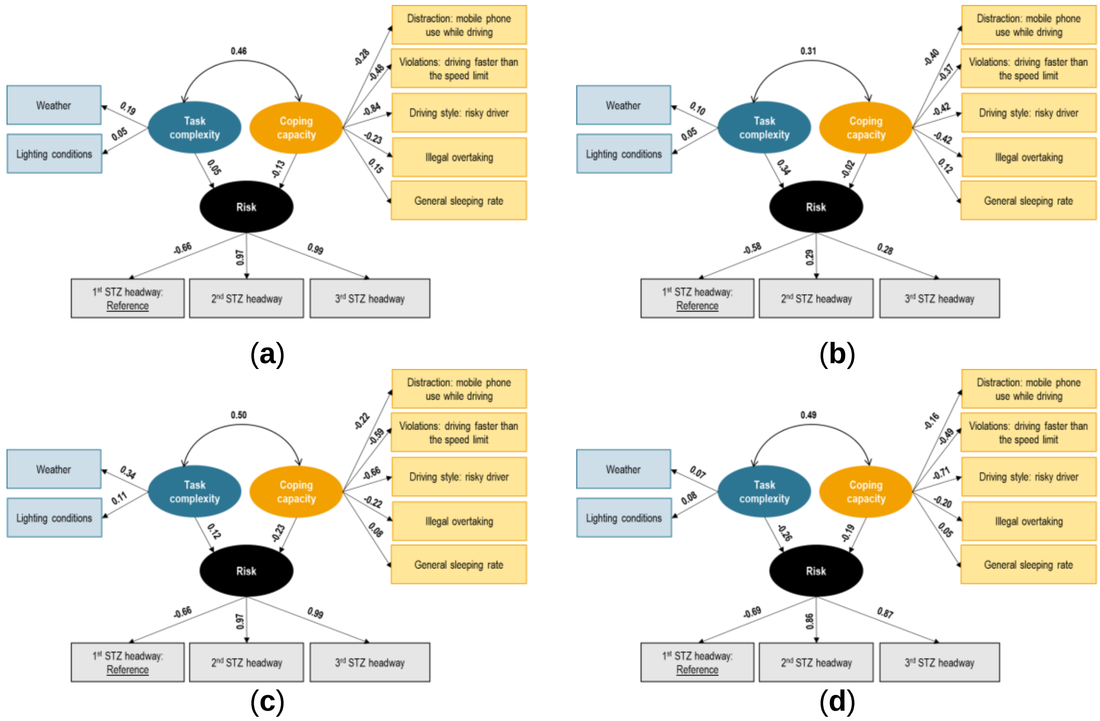

Figure 5: Results of SEM on risk – UK car drivers – experiment phase 1 (a), 2 (b), 3 (c), 4 (d).

5.2.3. German cars

To begin with, the risk is measured by means of the STZ levels for speeding (level 1 'normal driving' used as the reference case; level 2 refers to 'dangerous driving', while no incidents with regards to level 3 'avoidable accident driving' were found).

The structural model between the latent variables shows some interesting findings: first, task complexity and coping capacity are interrelated with a positive correlation (regression coefficient=0.003) – which reduces in magnitude as the drivers progress from phases 1 and 2 through phases 3 and 4. This positive correlation indicates that higher task complexity is associated with higher coping capacity implying that drivers' coping capacity increases as the complexity of driving tasks increases. Overall, the structural model between task complexity and risk shows a positive coefficient, which means that increased task complexity relates to increased risk according to the model (regression coefficient=8.11). On the other hand, the structural model between coping capacity and risk shows a negative coefficient, which means that increased coping capacity relates to decreased risk according to the model (regression coefficient=-0.25).

It is identified that the measurement equations of task complexity and coping capacity are consistent between the different phases. At the same time, the loadings of the observed proportions of the STZ of speeding are consistent between the different phases. The structural model between task complexity and inverse risk (normal driving) are positively correlated among the four phases while coping capacity and risk were found to have a negative relationship in all phases of the experiment.

In Germany, the model for speeding revealed a positive correlation between task complexity and coping capacity, but with the largest correlation in phase 2 of the experiment, where real-time warnings were introduced. At the end of the experiment (phase 4), coping capacity was found to have its largest correlation with risk, while task complexity had its greatest loading during phase 3 of the experiment. The results for all phases are shown in Figure 6 below.

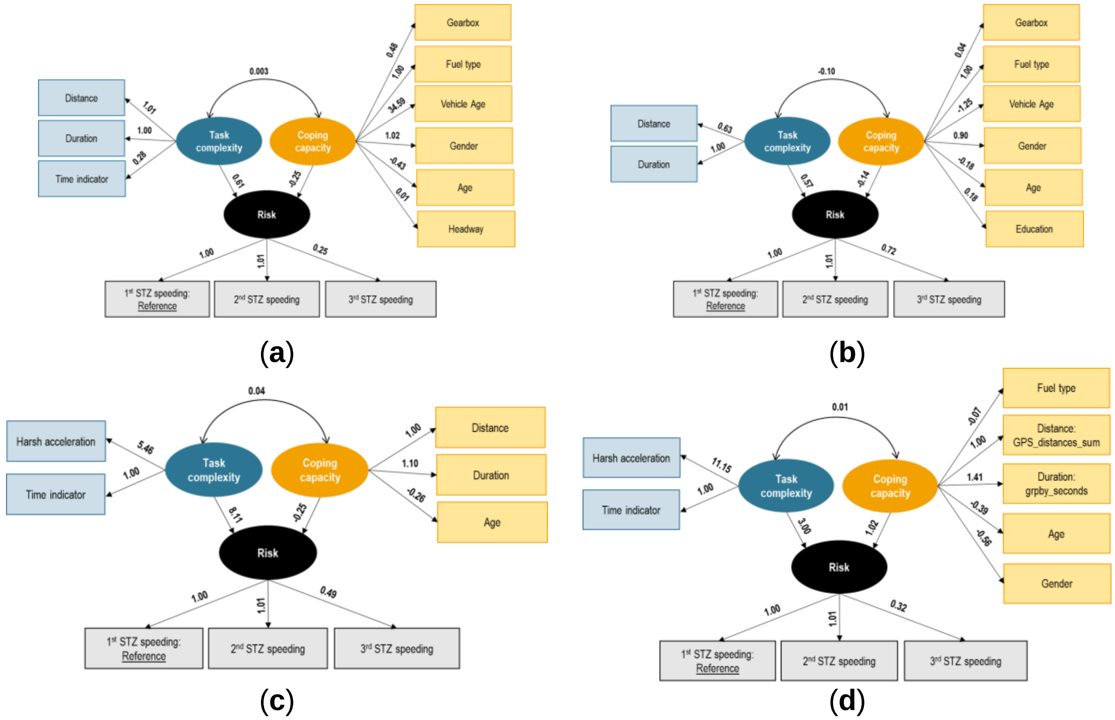

Figure 6: Results of SEM on risk – German car drivers – experiment phase 1 (a), 2 (b), 3 (c), 4 (d).

Table 5 summarizes the model fit of SEM applied for different counties (i.e. Belgium, UK, Germany) and experimental phases.

Table 5. Model Fit Summary for different counties, transport modes, and experimental phases

Model Fit measures	Phase 1	Phase 2	Phase 3	Phase 4
Belgian Cars				
AIC	273200.6	57294.26	338636.6	271111.2
BIC	273402.4	57518.77	338808.6	271253.0
CFI	0.661	0.473	0.484	0.817
TLI	0.560	0.335	0.291	0.709
RMSEA	0.121	0.082	0.103	0.037
UK Cars				
AIC	6377.390	4939.518	5266.238	7536.846
BIC	6599.142	5171.580	5489.058	7770.156
CFI	0.984	0.885	0.988	0.989
TLI	0.977	0.834	0.983	0.985
RMSEA	0.042	0.037	0.037	0.035
Germany Cars				
AIC	813827.574	676463.527	282420.347	525983.888
BIC	814118.257	676746.197	282625.175	526243.996
CFI	0.981	0.960	0.996	0.978
TLI	0.974	0.944	0.993	0.966
RMSEA	0.079	0.117	0.059	0.100

6. Discussion

Through the application of SEM models, the analyses revealed that higher task complexity levels lead to higher coping capacity by drivers. Additionally, the effect of task complexity on risk was greater than the impact of coping capacity in Belgium and Germany, while mixed results were observed in the UK. Models fitted on data from different phases of the experiments validated that interventions had a positive influence on risk compensation, increasing drivers' coping capacity and reducing dangerous driving behaviour.

As task complexity increased, drivers may experience greater cognitive load and divided attention, potentially leading to decreased situational awareness and slower response times. These factors can impair decision-making abilities and increase the likelihood of errors or collisions.

Higher task complexity was associated with an increased crash risk due to several reasons. Firstly, drivers could probably become overwhelmed by the demands of complex tasks, leading to reduced attention to the road and other traffic participants. This can result in delayed detection of critical events and inadequate responses. Secondly, complex tasks may require drivers to allocate more mental resources, causing them to divert attention from essential driving activities. For instance, interacting with in-vehicle technology or navigation systems can increase cognitive workload and lead to decreased focus on the primary task of driving.

Conversely, drivers with limited coping capacity may struggle to effectively manage complex tasks, leading to higher crash risk. Reduced coping capacity can manifest as slower reaction times, impaired judgment, and difficulties in prioritizing information. In situations where the demands of the driving task exceed a driver's coping capacity, there is an increased likelihood of errors, misjudgements, and collisions.

It is worth noting that the relationship between task complexity and risk, as well as coping capacity and risk, may depend on the specific context and the type of task or activity involved. In general, higher task complexity may increase the potential for errors or crashes, as it can lead to greater cognitive or physical demands on the individual performing the task. However, it is also possible that increased experience or training can help to mitigate the risk associated with higher task complexity. Similarly, a higher coping capacity may help to reduce the risk of crashes or errors, as it can provide individuals with the resources or strategies needed to effectively manage challenging or stressful situations. However, the effectiveness of coping strategies may depend on the specific context and the individual's ability to apply them in real-world situations. Overall, it is important to consider the specific factors and context involved when assessing the relationship between task complexity, coping capacity, and risk.

The developed models presented in this work can be further exploited by researchers and practitioners. Additional task complexity and coping capacity factors, such as road type, more personality traits and driving profiles could be utilized for example. Furthermore, data could be enhanced by including additional measurements such as electrocardiogram and electroencephalogram readings, traffic conflicts and transport emissions. Finally, additional methodologies such as imbalanced learning and models taking into account unobserved heterogeneity could be explored for the understanding of the relationship between task complexity, coping capacity and crash risk.

7. Conclusions

The objective of the present research was to model the inter-relationship between driving task complexity, coping capacity and crash risk using the i-DREAMS database. For that purpose, data collected from a naturalistic driving experiment with a sample of 133 drivers were utilized and data from Belgian, German and UK car drivers were collected and analysed. Explanatory variables of risk and the most reliable indicators, such as time headway, distance travelled, speed, forward collision, time of the day (lighting indicators) or weather conditions were assessed.

Results showed that higher task complexity levels lead to higher coping capacity. This means that drivers, when faced with difficult conditions, tend to regulate well their capacity to apprehend potential difficulties, while driving. It was revealed that the SEM applied between task complexity and inverse risk were positively correlated in all phases of the experiment, which means that increased task complexity relates to increased risk. On the other hand, coping capacity and inverse risk found to have a negative relationship in all phases, which means that increased coping capacity relates to decreased risk. Overall, the interventions had a positive influence on risk, increasing the coping capacity of the operators and reducing the risk of dangerous driving behaviour.

The integrated treatment of task complexity, coping capacity and risk can improve behaviour and safety of all travellers, through the unobtrusive and seamless monitoring of behaviour. Thus, authorities may use data systems at population level to plan mobility and safety interventions, set up road user incentives, optimize enforcement and enhance community building on safe traveling.

All in all, the inter-relationship between driving task complexity, coping capacity, and crash risk is a multifaceted and crucial area of study in traffic safety research. Driving task complexity refers to the level of demand and cognitive load imposed on the driver by various factors such as traffic density, road conditions, weather, and the presence of distractions. Coping capacity, on the other hand, encompasses the individual driver's ability to effectively manage and adapt to these complex driving tasks. It includes factors like driver experience, skills, perceptual abilities, decision-making processes, and the availability of appropriate coping strategies. The interplay between driving task complexity and coping capacity directly have a direct impact on crash risk, as drivers who are overwhelmed by high task complexity and have limited coping capacity may experience reduced situational awareness, slower reaction times, impaired decision-making, and increased likelihood of errors or collisions. Conversely, drivers with better coping capacity can effectively handle complex driving tasks, mitigate risks, and maintain safer driving behaviours.

Abbreviations

AIC: Akaike Information Criterion
BIC: Bayesian Information Criterion
CFI: Comparative Fit Index
FCW: Forward Collision Warning
GLMs: Generalized Linear Models
LDW: Lane Departure Warning
PCW: Pedestrian Collision Warning
RMSEA: Root Mean Square Error Approximation
SEMs: Structural Equation Models
STZ: Safety Tolerance Zone
TLI: Tucker-Lewis Index
VIF: Variance Inflation Factor

Acknowledgments
The research was funded by the European Union's Horizon 2020 i-DREAMS project (Project Number: 814761) funded by European Commission under the MG-2-1-2018 Research and Innovation Action (RIA).

Funding
The research was funded by the European Union's Horizon 2020 i-DREAMS project (Project Number: 814761) funded by European Commission under the MG-2-1-2018 Research and Innovation Action (RIA).

Availability of data and materials
Not applicable.

References

Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. Structural Equation Modeling: A Multidisciplinary Journal, 14(3), 464–504.
Collins, M., Dasgupta, S., & Schapire, R. E. (2001). A Generalization of Principal Component Analysis to the Exponential Family. Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic, 617–624.
Harrison, L., Stephan, K., & Friston, K. (2007). Effective Connectivity. In Statistical Parametric Mapping (pp. 508–521).
Hastie, T., & Tibshirani, R. (1990). Exploring the Nature of Covariate Effects in the Proportional Hazards Model. Biometrics, 46(4), 1005.
Hastie, T. J., & Pregibon, D. (2017). Generalized linear models. In Statistical models in S (pp. 195–247). Routledge.
Shah, S., Ahmad, N., Shen, Y., Pirdavani, A., Basheer, M., & Brijs, T. (2018). Road Safety Risk Assessment: An Analysis of Transport Policy and Management for Low-, Middle-, and High-Income Asian Countries. Sustainability, 10(2), 389.
Washington, S., Karlaftis, M., Mannering, F., & Anastasopoulos, P. (2020). Statistical and Econometric Methods for Transportation Data Analysis. Chapman and Hall/CRC.
Yuan, K.-H., & Bentler, P. M. (2006). 10 Structural Equation Modeling (pp. 297–358).