Gert Weller

The Psycho
Driving on

Development an
of a Model

Gert Weller

The Psychology of Driving on Rural Roads

VS RESEARCH

Verkehrspsychologie

Herausgegeben von
Prof. Dr. Bernhard Schlag, TU Dresden

Verkehrspsychologie ist ein wachsendes Gebiet der Psychologie, das starke öffentliche Aufmerksamkeit erfährt. Empirische Forschung in der Verkehrspsychologie umfasst neben der Diagnostik und Rehabilitation auffälliger Kraftfahrer eine Reihe innovativer Gebiete, deren gemeinsamer Erkenntnisgegenstand das Mobilitätsverhalten und Mobilitätserleben der Menschen ist. Verkehrspsychologische Forschung wird oft in enger Kooperation mit Ingenieuren, Wirtschaftswissenschaftlern und Medizinern betrieben und hat dabei teilweise eigenständige theoretische und methodische Ansätze entwickelt. Die Bände dieser Reihe befassen sich u. a. mit dem Mobilitätsverhalten und der Verkehrsmittelwahl, Möglichkeiten der Verhaltensbeeinflussung für eine umweltgerechtere und sicherere Mobilität, psychologischen Aspekten der Verkehrsplanung und des Mobilitätsmanagements, Fragen der Unfallforschung und der Verbesserung der Verkehrssicherheit, der Fahrerassistenz sowie der Akzeptanz von und dem Umgang mit technischen und organisatorischen Innovationen. Die Reihe macht sowohl aktuelle Forschungen als auch Überblicksdarstellungen in diesen Bereichen zugänglich.

Gert Weller

The Psychology of Driving on Rural Roads

Development and Testing of a Model

VS RESEARCH

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

Dissertation Technische Universität Dresden, 2009

1st Edition 2010

Editorial Office: Verena Metzger / Britta Göhrisch-Radmacher

VS Verlag für Sozialwissenschaften is a brand of Springer Fachmedien.
Springer Fachmedien is part of Springer Science+Business Media.
www.vs-verlag.de

Cover design: KünkelLopka Medienentwicklung, Heidelberg
Printed on acid-free paper
Printed in Germany

ISBN 978-3-531-17529-4

Acknowledgements

Writing this thesis was made possible by the support of many individuals and institutions.

I would like to thank my supervisor Prof. Dr. Bernhard Schlag who not only provided guidance and valuable advice but also encouraged and supported the development of my own ideas. Thank you for your trust and patience!

Thank you also to all my colleagues on the RIPCORD-iSEREST project and especially the ones with whom I worked on specific sub-tasks: it was a pleasure to work, discuss and travel together with you, the last both literally and figuratively. I am also indebted to the European Commission for financing RIPCORD-iSEREST during which the empirical research reported in this thesis was conducted.

Many thanks to all my present and former colleagues at the Chair of Traffic and Transportation Psychology; your ideas and encouragement were greatly appreciated.

Special thanks to Dr. Dirk Ebersbach, Matthias Dietze and all other members of the Chair of Road Planning and Road Design; your generous help and professional expertise allowed the field study to be conducted.

I am also grateful to the former students who worked on the project and largely contributed to this thesis. Special thanks to the ones who wrote their diploma thesis on the project: Tino Friedel, Ina Petermann and Jana Voigt.

Many thanks to Dr. Thoralf Knote, Dr. Günther Nirschl and Marina Voigtländer from the Fraunhofer IVI in Dresden for your friendly assistance. The simulator study reported in this thesis was made possible by your support.

I would also like to thank all the participants in the various studies. This is especially the case for all colleagues of the TU Dresden who participated in the field study although they probably had better things to do - thank you!

Many thanks to Prof. Dr. Winfried Hacker and Prof. Dr. Josef Krems for agreeing to be the expert reviewers of the thesis.

Gert Weller

Zusammenfassung auf Deutsch

Die vorliegende Arbeit wurde mit dem Ziel verfasst, einen Beitrag zur Erhöhung der Verkehrssicherheit auf Landstraßen zu leisten. Landstraßen sind die gefährlichste Straßenkategorie, wenn die Anzahl der Getöteten betrachtet wird (Statistisches Bundesamt, 2007). Dies ist auf die Spezifika dieser Straßenkategorie in Kombination mit dem dort gezeigten Fahrverhalten zurückzuführen (OECD, 1999). Damit ist eine Erhöhung der Verkehrssicherheit nur möglich, wenn gleichzeitig straßen- und situationsseitige Merkmale und deren Wirkung auf das menschliche Verhalten betrachtet werden.

Ausgangspunkt der Arbeit war zunächst die Zusammenfassung vorliegender Forschungsbefunde und die darauf aufbauende Entwicklung eines eigenen psychologischen Modells zum Fahren auf Landstraßen. Die Validierung des Modells erfolgte anhand einzelner daraus abgeleiteter Hypothesen in drei empirischen Untersuchungsschritten.

Im ersten Schritt der empirischen Validierung wurde untersucht, wie unterschiedliche Landstraßen und deren Wahrnehmung die dort zu fahrende angemessene Geschwindigkeit beeinflussen. Dazu wurden 21 Landstraßenphotos von 46 Probanden nach verschiedenen Merkmalen beurteilt. Die Beurteilung der Wahrnehmung erfolgte mit einer deutschen Version der von Steyvers (Steyvers, 1993, 1998; Steyvers, Dekker, Brookhuis, & Jackson, 1994) entwickelten Road Environment Construct List (RECL).

Voraussetzung einer reliablen Erfassung der Merkmale war die Übereinstimmung der Faktorenstruktur der deutschen Version der RECL mit der Originalversion. Bei dieser vorangestellten Überprüfung zeigten sich jedoch deutliche Unterschiede. Mögliche Ursachen dafür werden im Text diskutiert, wobei auch auf die statistischen Hintergründe faktorenanalytischer Methoden eingegangen wird. Aufgrund dieser Ergebnisse habe ich mich entschlossen die Analyse von Wahrnehmungsaspekten nachfolgend auf zwei „Marker-Items“ zu beschränken. Die Auswahl der Marker-Items „gefährlich“ und „monoton“ erfolgte theoriebasiert und anhand statistischer Kenngrößen. Eine regressionsanalytische Überprüfung des Einflusses der Marker-Items auf die beurteilte angemessene Geschwindigkeit ergab hochsignifikante Einflüsse. Die gesamte Varianzaufklärung war zudem nicht wesentlich geringer als die ursprünglichen Berechnungen mit aus allen Items abgeleiteten Faktoren (Weller, Schlag, Friedel, & Rammin, 2008).

In einem weiteren Auswertungsschritt dieser im Labor erhobenen Daten wurde mit einfachen strukturanalytischen Modellen überprüft, inwieweit differentielle Faktoren bei der Beurteilung der Landstraßen und der angemessenen Geschwindigkeit beteiligt sind. Diese Frage war nicht nur für das Modell relevant, sondern hätte abhängig von den Ergebnissen auch Konsequenzen für die Gestaltung von Landstraßen. Unter Verwendung der Organismusvariable „Alter“ konnte für eine ausgewählte Stichprobe von Landstraßen, neben dem indirekten Einfluss über die Wahrnehmung, zusätzlich ein direkter Einfluss auf die Geschwindigkeit nachgewiesen werden (partial mediation model, James, Mulaik, & Brett, 2006). Wegen des statistisch relativ schwachen Befundes und der Beschränkung auf ausgewählte Landstraßen erfolgte zu diesem Zeitpunkt keine Integration der Befunde in das bestehende Modell.

Im zweiten empirischen Untersuchungsschritt wurden Fahrversuche im Simulator des Fraunhofer IVI durchgeführt (N = 50). Ziel dieses Erhebungsschrittes war die experimentelle Untersuchung des Einflusses einzelner Gestaltungsmerkmale von Kurven und geraden Streckenabschnitten. Neben den abhängigen Variablen Wahrnehmung und Erwartung hinsichtlich eines angemessenen Verhaltens war es nun auch möglich tatsächlich gefahrene Geschwindigkeiten (wenn auch simuliert) in die Modellvalidierung zu integrieren. Darüber hinaus erlaubte das experimentelle Design die Formulierung und Überprüfung von Hypothesen.

Vor der eigentlichen Datenauswertung musste sichergestellt werden, dass die im Simulator erhobenen Daten als Proxy für reales Verhalten geeignet waren. Dies erfolgte zunächst über vorangestellte Auswertungen zur Überprüfung des Einflusses verschiedener Störvariablen. Hauptsächlich gelang der vorläufige Nachweis der externen Validität in einer Analyse der Daten unter Berücksichtigung von Geschwindigkeitsprognosemodellen nach Lamm et al. (2007).

Die eigentliche hypothesengeleitete Analyse der Daten erbrachte einen hoch signifikanten Einfluss sowohl der Umfeldgestaltung von Geraden als auch einzelner Gestaltungselemente im Sinne von Cues (Posner, Snyder, & Davidson, 1980) oder Signalen (Hacker, 2005) bei Kurven. Von besonderem Interesse ist hier das Ergebnis, dass eine Geschwindigkeitsverringerung nicht nur durch formale Warnschilder erfolgte, sondern auch durch eine Verringerung der Sichtdistanz in der Kurve.

Für die im Modell angenommene Wirkungskette, ausgehend von objektiven Merkmalen, deren bewusster Wahrnehmung und daraus abgeleiteten Erwartungen hin zum tatsächlichen Verhalten, ergaben sich unterschiedliche Befunde für Kurven und Geraden. Während die Annahmen für Kurven gut bestätigt werden konnten, war das Geschwindigkeitsverhalten auf Geraden nicht durch die bewusste Wahrnehmung, erhoben mit Ratings, zu erklären. Hingegen lieferte der optische Fluss nach Gibson (1986) den theoretischen Hintergrund zur Erklärung der Resultate für Geraden.

Entsprechend diesen Ergebnissen erfolgt eine Erweiterung des bestehenden Modells um einen zweiten direkten Einfluss von Streckenmerkmalen auf das Verhalten.

Die Verhaltenssteuerung über diesen Pfad erfolgt weitestgehend über den optischen Fluss und perzeptuelle Invarianten wie time-to-collision oder tau (Gibson, 1986). Dieser Pfad kann wegen seiner ständigen optischen Rückmeldungen als closed-loop control bezeichnet werden. Er steht damit im Gegensatz zum Voraus gerichteten indirekt gesteuerten Pfad der einen open-loop darstellt. Über diesen Pfad werden Verhaltensanpassungen gesteuert, die vor der eigentlichen Situation und deswegen ohne direkte Rückmeldung erfolgen.

Als Kontrollinstanz zur Auswahl eines der beiden Pfade zur Steuerung des Verhaltens wird ein Aufmerksamkeitsmonitor angenommen, ähnlich den Aufmerksamkeits-Checks im GEMS von Reason (1990) oder dem subjektiven Risiko Monitor im Modell von Näätänen & Summula (1974). Über diesen Aufmerksamkeitsmonitor wird entschieden, ob sich die voraus liegende Situation von der aktuellen Situation unterscheidet oder unerwartete Ereignisse eine Anpassung des Verhalts erforderlich machen.

Ähnlich den Laborversuchen erfolgte auch in den Simulatorversuchen eine Überprüfung des Einflusses von Personenmerkmalen auf das Verhalten. Während der Einfluss von Organismusvariablen bei der Beurteilung der im Labor erhobenen Geschwindigkeitserwartungen noch teilweise über die Wahrnehmung der Streckenmerkmale erfolgte, wurde für die im Simulator erhobenen Daten nur noch ein direkter Einfluss auf das Verhalten gefunden. Damit war dieser Einfluss erst zu einem späten Zeitpunkt wirksam und muss als Kalibrierungsfaktor nachfolgend der Wahrnehmung verstanden werden. Wie bei der Erweiterung des Modells um den zweiten Pfad, wird das Modell um den direkten Einfluss von Personenmerkmalen auf das Verhalten ergänzt. Da diese Erweiterungen post-hoc erfolgten, bedürfen sie der experimentellen Validierung in zukünftigen Versuchen.

Der dritte empirische Erhebungsschritt im Rahmen dieser Arbeit befasste sich mit dem Einfluss des Verhaltens auf Unfälle. Dazu wurden mit 16 Probanden Fahrversuche im Feld mit dem Messfahrzeug des Lehrstuhls für Straßenplanung der TU Dresden durchgeführt. Im Modell wird angenommen, dass es zu Unfällen kommt, sobald das tatsächliche Verhalten vom angemessenen Verhalten abweicht. Im hier verwendeten Untersuchungsparadigma wurden Kurven mit einer hohen Unfallrate mit geometrisch ähnlichen Kurven, jedoch geringerer Unfallrate verglichen. Das gemessene Verhalten in den Kurven mit niedriger Unfallrate wurde als angemessenes Verhalten definiert und diente als Referenz für das Verhalten in Kurven mit hoher Unfallrate. Auf der gefahrenen Strecke nördlich von Dresden lagen vier Kurven mit hoher Unfallrate. Der Kurs wurde sowohl in Hin- als auch in Rückrichtung gefahren und erstreckte sich damit auf insgesamt etwa 80 Kilometer.

Unter Rückgriff auf Homöostasemodelle des Fahrverhaltens (Fuller, 2005; Fuller, McHugh, & Pender, 2008; Wilde, 1988, 2001) wurde Verhalten nicht nur über die Geschwindigkeit definiert, sondern auch über die Beanspruchung und die Blickbewegung. Im vorliegenden Fall wurde die Beanspruchung über die Reaktionszeiten auf eine visuelle Nebenaufgabe erhoben. Das Blickverhalten wurde mit Hilfe des im Messfahrzeug integrierten berührungslosen Messsystem Smart Eye ermittelt.

Hypothetisch angenommene Unterschiede im Verhalten ließen sich nicht finden. Erklärt wird dies mit unterschiedlichen Modellannahmen der Unfallentstehung. Die Versuche wurden unter der Annahme durchgeführt, dass Unfälle durch eine Veränderung des durchschnittlichen Verhaltens in Richtung kritisches Verhalten entstehen. Dies war zumindest für die untersuchten Kurven nicht der Fall. Dort ist offenbar eine deutliche Abweichung des Verhaltens weniger Fahrer die Unfallursache, nicht die Veränderung des Durchschnittsverhaltens. Eine empirische Bestätigung der Modellannahmen hinsichtlich Unfallentstehung konnte damit nicht erbracht werden.

Eine indirekte Validierung der Annahmen zur Unfallentstehung erfolgte mit einer zusätzlichen Auswertung des Einflusses der Kurvigkeit von Einzelkurven auf die Beanspruchung und die Geschwindigkeit. Mit zunehmender Kurvigkeit zeigte sich sowohl eine Abnahme der Geschwindigkeit als auch eine Zunahme der Beanspruchung. Das gleichzeitige Auftreten beider Befunde spricht gegen die Annahme homöostatischer Prozesse beim Kurvenfahren. Die bekannten Zusammenhänge zwischen Kurvigkeit und Unfallgeschehen (Elvik & Vaa, 2004) in Kombination mit diesen Ergebnissen lassen vermuten, dass der im Modell angenommene Zusammenhang zwischen Verhalten und Unfällen besteht, auch wenn er in der vorangegangenen Auswertung nicht nachgewiesen werden konnte.

Unabhängig von dem Modell wurde in einer weiteren zusätzlichen Auswertung untersucht, wie sich die Bearbeitung der Zweitaufgabe auf das Blickverhalten auswirkt. Es zeigte sich, dass das Blickverhalten eine eindeutige Zuordnung, ob mit oder ohne Nebenaufgabe gefahren wurde, ermöglicht. Dieses Ergebnis ist vor allem für die Entwicklung zukünftiger Fahrerinformations- und Fahrerassistenzsysteme relevant.

Zusammenfassend kann festgehalten werden, dass mit dem Nachweis zweier Wirkpfade der Verhaltenssteuerung beim Fahren auf Landstraßen eine gezielte Verhaltensbeeinflussung über die optische Gestaltung des Umfeldes und der Streckenmerkmale möglich ist. Damit ergeben sich weitgehende Möglichkeiten zur Reduzierung der Unfallzahlen, was mit dem eingangs formulierten Ziel dieser Arbeit in Einklang steht.

Abstract in English

This thesis was written as a contribution to increasing traffic safety on rural roads. These roads constitute the most dangerous road category when the number of fatal accidents is taken into account (Statistisches Bundesamt, 2009). Their dangerous nature can be attributed to their inherent properties in combination with the driving behaviour that occurs on them (OECD, 1999). Thus, increasing traffic safety on these roads requires not only taking into account the characteristics of the road and the situation but also their effect on human behaviour.

The starting point of the work was the synopsis of existing research and the subsequent development of a psychological model for driving on rural roads. The validation of this model was conducted in three empirical steps based on hypotheses derived from the model.

In the first step, the influence of various rural roads and perception thereof on perceived appropriate speed was analysed. Forty-six participants rated 21 pictures of rural roads along various dimensions. For the ratings, a German version of the Road Environment Construct List (RECL) developed by Steyvers (Steyvers, 1993, 1998; Steyvers et al., 1994) was used.

For a reliable assessment of the road characteristics it was important to ensure that the factor structure of the original RECL and of the German version of the RECL corresponded. However, this preliminary assessment yielded several differences. Potential reasons for this finding are discussed in the text whereby the statistical background of factor analysis is taken into account. Subsequent to the results a decision was made to limit further analyses to two marker items. The selection of the items “dangerous” and “monotonous” was based on theoretical and statistical indicators. A regression of the ratings of these two marker items on rated speed revealed a highly significant influence. The explained variance of this regression was comparable to the original calculation which encompassed all items (Weller, Schlag et al., 2008).

Structural equation models were performed for additional analyses of the laboratory data. This was done to test how differential characteristics of the participants influenced the perception of the roads and the ratings of appropriate speed. This question was not only relevant for the development of the model but also has consequences for the design of rural roads. For the variable “age”, a direct influence on speed was found in addition to the indirect influence on speed due to

perception (Partial Mediation Model, James et al., 2006). Integration of these results into the model was not performed at this stage because of the relatively weak results and the restriction to a subset of the roads.

In the second empirical step driving experiments were conducted in the simulator of the Fraunhofer IVI (N = 50). The purpose of this step was to test the influence of various design variants of both single curves and straight road sections. The experimental nature of this research step allowed the formulation and testing of hypotheses. Besides the two dependent variables of "perception" and "expectations regarding appropriate speed", it was now also possible to incorporate actual driven speeds – albeit in a simulated environment - into the model evaluation.

Prior to the actual data analyses it was important to ensure that the data collected in the simulation could be used as a proxy for real behaviour in the field. Initially, this was done by analysing the potential influence of confounding variables. Subsequently, preliminary proof of the external validity of the data was achieved by using speed prediction models developed by Lamm et al. (2007).

The subsequent hypotheses-driven analyses of the data revealed a highly significant influence from both the environmental characteristics of straight road sections and from single design elements which served as cues (Posner et al., 1980) or signals (Hacker, 2005) in curves. Of special interest for road planning and road safety is the finding that a reduction in speed was not only found for curve warning signs but also due to a reduction in sight distance in the curves.

The results for curves and straight road sections differed when it came to the assumed chain of events proposed in the model. While the model assumptions were met for curves, behaviour on straight road sections could not be explained by conscious perception as defined by the ratings. For straight road sections, the optic flow (Gibson, 1986) provided the theoretical background for explaining behaviour.

Based on these results, a second direct influence of objective road characteristics on behaviour was implemented in the model. Via this path, behaviour is mainly controlled by the optic flow and perceptual invariants such as time-to-collision and tau (Gibson, 1986). Due to this constant feedback, this path constitutes a closed-loop-control. It thus contrasts to the open-loop control used for an anticipated adaptation of behaviour. This adaptation takes place before the actual situation is encountered and is thus implemented without feedback.

The selection of one of the two paths assumes the existence of an attention monitor. This control authority is seen as being similar to the attentional checks in the GEMS Model of Reason (1990) or the subjective risk monitor in the model of Näätänen & Summala (1974). The decision within the attention monitor is based on whether the upcoming situation differs from the current situation or whether unexpected events require an adaptation of behaviour.

The simulator study was similar to the laboratory study in that it tested whether individual variables influenced behaviour. For the simulator study only a direct influence of these variables on behaviour was found. An indirect influence via the perception of the road characteristics as was found in the laboratory study was not present. Thus, individual variables only affected behaviour at a rather late stage. Their influence must be understood as calibration of behaviour subsequent to the perception of the situation. As was done with the second path, the model was enlarged for this direct influence of individual variables on behaviour. These latter model enlargements require further experimental validation because they were conducted post-hoc.

The third empirical step dealt with the influence of behaviour on accidents. For this purpose, on-the-road driving experiments were conducted with 16 participants who drove an experimental vehicle. In the model it is assumed that accidents occur when actual behaviour deviates from appropriate behaviour. In the experimental paradigm used for these experiments, curves with a high accident rate were compared to curves with similar geometry but a low accident rate. The behaviour measured in the curves with the low accident rate was defined as appropriate behaviour and served as a reference for the behaviour measured in the high-accident rate curves. The experimental driving course contained four high-accident rate curves; it was driven in the outbound and inbound direction and in total amounted to 80 kilometres.

By referring to homeostatic models of driving behaviour (Fuller, 2005; Fuller et al., 2008; Wilde, 1988, 2001), the definition of behaviour was not limited to speed but also comprised workload and gaze behaviour. For this thesis, workload was assessed via the reaction time to a visual secondary task. Gaze behaviour was recorded via a contact-free eye-tracking system which was integrated into the measurement vehicle.

Differences in behaviour as were hypothetically assumed were not found. This is explained by conflicting model assumptions regarding accident occurrence: the driving experiments were conducted based on the assumption that accidents occur when average behaviour changes into more critical behaviour. At least for the experimental curves this was not the case. For these, rather than a change in average behaviour, a marked deviation in the behaviour of single individuals was presumably the reason for accident occurrence. Thus, the model assumptions regarding accident occurrence could not be validated empirically.

However, the model assumptions were validated indirectly with additional analysis of the influence of curvature of single curves on workload and speed. With increasing curvature, both a decrease in speed and an increase in workload were found. This concurrence is evidence against homeostatic processes in curve driving. The combination of these findings and the known relationship between curvature and accident occurrence (Elvik, Høye, Vaa, & Sørensen, 2009) suggests that the assumed relationship between behaviour and accidents exists although it could not be shown in the preceding analyses.

In addition to testing the model assumptions it was analysed how attending to the secondary task changed gaze behaviour. It was shown that gaze behaviour was a reliable indicator of whether the driver attended to the secondary task or not. This finding is highly relevant for the development of future driver assistance and information systems.

In summary, the results provide a starting point for the design of behaviourally relevant roads and road environments. This can contribute to a reduction of rural road accidents and is thus in line with the aims stated at the beginning of this thesis.

Contents

List of Tables

List of Figures

List of Abbreviations

-2LL	log likelihood times -2

β	beta, standardised regression coefficient
χ^2	chi-square test statistics
η^2	eta-squared
τ	Kendall's tau
$\tau(t)$	perceptual invariant tau
$\dot{\tau}$	tau dot

AADT	Average annual daily traffic
a	acceleration (in formulas)
acc	acceleration (abbreviated in tables)
ACC	adaptive cruise control
ADAS	advanced driver assistance systems
ADT	average daily traffic
ALM	allgemeines lineares Modell (English: GLM)
AMF	accident modification factor
ANOVA	analysis of variance
B	unstandardised regression coefficient
CAN	controller area network
CCR	curvature change rate
CCR$_S$	curvature change rate of the single curve
CI	confidence interval
C.R.	critical ratio
d	distance
D-B	Durbin-Watson
dec	deceleration
df	degree of freedom
DIN	Deutsche Industrie Norm
EDA	electrodermal activity
EDR	electrodermal response
e.g.	*exempli gratia (*for example)

F	F-ratio test statistics
FGSV	Forschungsgesellschaft für Straßen- und Verkehrswesen
ft	feet (US; multiply by 0.305 to find meter)
GIDAS	German in-depth accident study
GLM	generalised linear model (German: ALM)
GPS	global positioning system
GSR	galvanic skin response
GUI	graphical user interface
HMD	head-mounted display
HTML	hypertext markup language
Hz	Herz
i.e.	*id est* (that is)
IHSDM	Interactive Highway Safety Design Model
IOR	inhibition-of-return (paradigm)
ISO	International Standardization Organization
IVI	Institut für Verkehrs- und Infrastruktursysteme
IVIS	in-vehicle information systems
km/h	kilometre per hour
KMO	Kaiser-Meyer-Olkin (criterion)
K-S test	Kolmogorov-Smirnov test
LL	log likelihood
M	mean (arithmetic average)
m	metre
max.	maximum
MDS	multidimensional scaling
min.	minimum
ML	maximum likelihood
ms	millisecond
MS	mean square
MSA	measure of sampling adequacy
MW	Mann-Whitney test
n	number in a subsample
N	total number in a sample
n.s.	not significant
NS.	non-specific (EDR or SCR)
OAS	Operative Abbildsysteme
OECD	Organisation for Economic Cooperation and Development
OR	odds ratio
p	error probability
p.	page

PAF	principal axis factor analysis
PCA	principal component analysis
PDT	peripheral detection task
pp.	pages
PRP	psychological refractory period
r	Pearson product-moment correlation
R^2	explained variance (=r squared)
RECL	road environment construct list
RHT	risk homeostasis theory
s	second
SCL	skin conductance level
SCR	skin conductance response
SD	standard deviation
SDI	spatial density index
SE	standard error
SEM	structural equation modeling
SEU	subjective expected utility
SOA	stimulus onset asynchrony
SQL	standard query language
SS	sum of squares
SS_M	model sum of squares
SS_T	total sum of squares
StVO	Straßenverkehrsordnung (German road traffic regulations)
t	student t-test statistic
TCI	task capability interface
TL	tangent length
TLC	time to line crossing
TTC	time to collision or time to contact
TU	Technische Universität (technical university)
TUD	TU Dresden
v	velocity
vs.	versus
V85	85th percentile of speed
VBA	visual basic for applications
VCR	video cassette recorder
VIMS	visually induced motion sickness

1 Driving on Rural Roads: The Current Situation

Rural roads are the most dangerous road class in terms of the number of fatalities. The usual proportion of fatalities on rural roads is approximately 60% compared to 10% for motorways and 30% for inner urban roads (IRTAD, 2007; SafetyNet, 2007). The proportion of fatalities on rural roads has even increased over the last 25 years which reflects the comparatively successful interventions for motorways and inner urban roads (OECD, 1999). The combination of these facts underlies the high priority with which safety on rural roads should be addressed in the near future if the aims set by organisations such as the European Commission (2001) are to be met.

At the same time inherent properties of rural roads make such interventions difficult and costly. Amongst these properties are

- the often historical roots and the fact that the geometry of rural roads often fails to meet current safety standards;
- the different functions these roads have to fulfil;
- the comparatively high speed limits and high speeds driven;
- the large variation of speeds driven, both within and between users; and
- the unforgiving roadsides.

These properties lead to characteristic accident patterns. Driving accidents on rural roads are defined as the result of the driver losing control of the vehicle without the influence of other vehicles (FGSV, 2001) and account for more than 50% of all fatal accidents (Figures for Germany, Statistisches Bundesamt, 2006). When driver error is taken into consideration, inappropriate speed is seen as the main cause in 30% of all accidents involving personal injuries on rural roads (Statistisches Bundesamt, 2006). In general, human factors are seen as a contributing factor in more than 90% of all accidents (Treat et al., 1977; Weller et al., 2006 for a discussion).

These facts reveal the large potential for reducing severe accidents on rural roads once targeted action against human error and traffic violations is undertaken. To do so successfully, the psychological mechanisms underlying human error must be understood. This work is restricted to driving on two-lane rural roads (no autobahn-like roads) and does not take into account crossroads or inner-urban sections of rural roads.

2 Applying Existing Models to Driving on Rural Roads

In this chapter, different models of driver and driving behaviour are introduced and discussed with relation to their relevance to driving on rural roads. Firstly, models are presented which provide a framework into which other theories can be integrated. Secondly, individual differences are discussed; and thirdly, motivational models are introduced. The information needed as the input for motivational models is derived from perception and as such, models and theories of perception are summarised in a later chapter.

2.1 A Framework

When developing a model of a task in a specific situation, the task as such should first be analysed from a broader perspective. This has been done in the past for the driving task and has resulted in a number of widely accepted general framework models. Such framework models are necessarily descriptive and do not provide details of how the different components interact. Typical descriptive models of the driving task are control loop models (Durth, 1974) (see chapter 2.3.3) or hierarchical models such as the models of Michon (1971, 1979, cited in Michon 1985) and Janssen (1979, cited in Michon, 1985). Donges (1982, cited in Donges, 1999) combined this hierarchical model with the performance levels and the respective behavioural determinants (knowledge, rule, skill-based) described by Rasmussen (1986) as shown in Figure 1.

Driving is seen here as a hierarchical problem-solving task that comprises three different levels which can be divided by the specific task requirements at each level, the time frame needed to carry them out and the cognitive processes involved. The left section in Figure 1 represents the different task levels proposed by Rasmussen while the right section represents the model by Michon.

The strategic or navigational level comprises all processes concerning trip decisions such as where to go, when to go, which roads to take and what modes of transport to use. Decisions at this level are rare and take the longest in comparison to the other levels. Due to their nature they are processed in a more or less aware mode but become habits in case of constant repetition.

At the manoeuvring level, decisions are made within seconds. Typical manoeuvres are overtaking, turning or gap acceptance. Behaviour at the manoeuvring level is influenced by both motivational and situational variables. Other terms used to describe the manoeuvring level are tactical or guidance level.

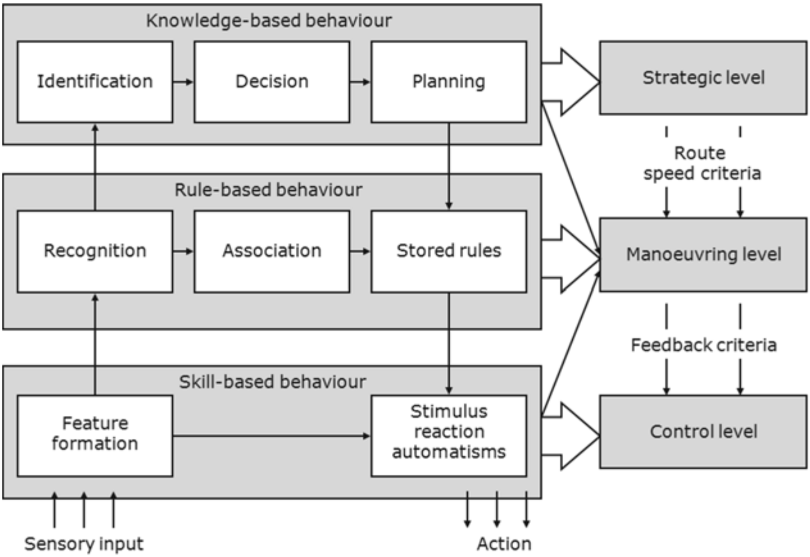


Figure 1. Combination of performance levels according to Rasmussen (1986) and the hierarchical model according to Michon (1985), modified from Donges (1982, in 1999).

Lastly, decisions at the control level are made rather automatically within a very short time range as stimulus response reactions. Typical tasks at this level are lane-keeping or gear shifting, although the latter example is not regarded as automatic by all authors, see Groeger (2000). For the control level, alternative terms such as operational or stabilisation level are used concurrently.

Whether a task is situated at the knowledge-based, rule-based or skill-based level depends to a great extent on the familiarity with the task and the environment and is also a function of driving experience. In general, higher order processes situated at the knowledge-based level require more cognitive resources than lower level processes. According to Schneider and Shiffrin (1977) and Shiffrin and Schneider (1977), higher levels of processing are usually referred to as controlled processing, whereas lower levels of processing are referred to as automatic processing.

A more detailed model of the principles of decision making and problem solving is depicted in Figure 2 (Reason, 1990). This model can also be applied to the context of driving as described above.

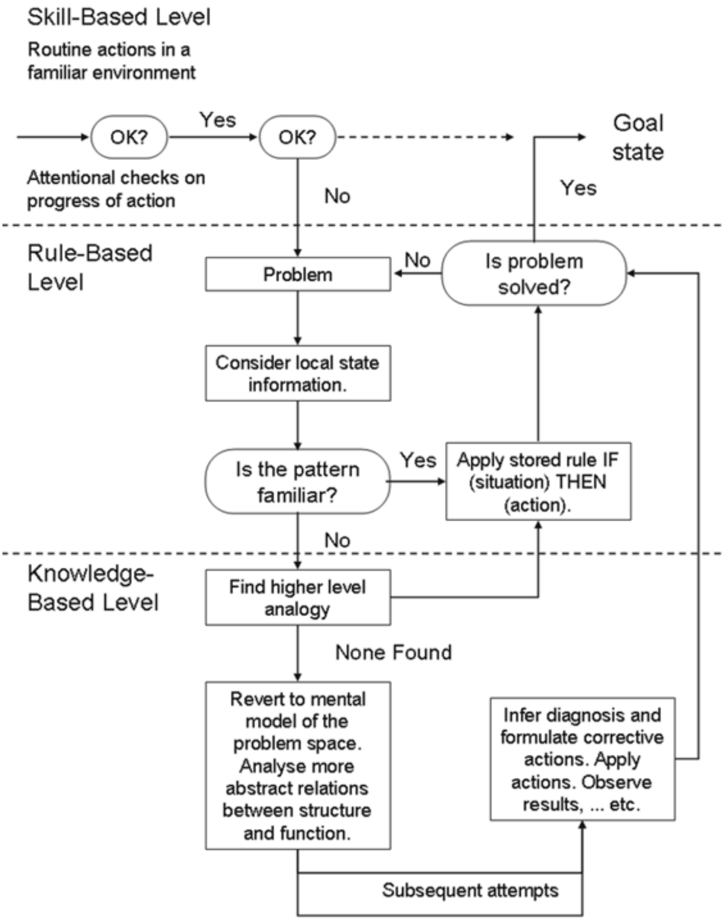


Figure 2. The generic error-modelling system (GEMS) as proposed by Reason, (1990).

The crucial point for rural road design is that people, in general, tend to rely on pre-programmed behavioural sequences found at the skill-based level than revert to higher-order processing. This is because the latter requires more resources (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). Similarly, rule-based behaviour is preferred to knowledge-based behaviour as '… humans, if given a choice, would prefer to act as context-specific pattern recognisers rather than attempting to calculate or optimise' (Rouse, 1981, cited in Reason, 1990, p. 65).

2.2 Individual Differences: Traits and Demographic Variables

The descriptive models introduced in the previous chapter were developed irrespective of different driver groups. This in itself constitutes a development in the history of psychological driver and driving behaviour modelling and broadens the initial approach which was exclusively based on differences between driver groups. Nevertheless, understanding the cognitive and motivational reasons resulting in these differences still provides important input for the development of a model. Furthermore, understanding of group specific differences is needed both in the planning phase of validation studies including the selection of participants as well as for the interpretation of the results following the studies.

A very early concept in traffic psychology was the notion of an accident-prone driver. Haight (2004) describes the historic rise and fall of the term. One reason why the concept has not been successful is related to the nature of accidents which are rare events from a statistical perspective (see Weller, Schlag, Gatti et al., 2006 for an overview). Due to these characteristics, predictions of future accident involvement for an individual driver based on past accident records lead to a high number of false positives or false negatives at this individual level (Evans, 2004; Klebelsberg, 1982).

However, when accident involvement is looked at from an aggregated level, it does indeed show characteristic differences between driver groups. Thus, the likelihood of being involved in an accident is assigned to different driver groups instead of predicting accident involvement for an individual. When developing such group specific functions of accident likelihood, the basic approach is to use available accident data and analyse it with respect to variables such as gender, age or driving experience.

The age related function of accident involvement per unit time or distance driven is U-shaped with the youngest and oldest drivers (aged 70 plus) showing the highest risk (Elvik & Vaa, 2004; Schade, 2000). With the increasing population of older drivers, this fact is highly relevant for traffic safety (overview in Schlag, 2008a; Schlag, 2008b). The reasons behind the higher accident involvement of younger and older drivers are different. For the younger age group motivational factors are more important (Boyce

& Geller, 2002; Chipman, MacGregor, Smiley, & Lee-Gosselin, 1993; OECD, 2006; Schlag, 1994), whereas for the older drivers the age related decrease in physiological and cognitive functioning can be seen as a reason (Ellinghaus, Schlag, & Steinbrecher, 1990; Schlag, 1993, 2008c; Weller & Geertsema, 2008).

This decline also affects driving behaviour. Here, two effects are reported in the literature (overview and discussion in Weller & Geertsema, 2008). First of all, a main effect of age is often reported whereby older drivers drive slower, have higher reaction times and drive with fewer acceleration/deceleration actions (Breker et al., 2002/2003; Chaparro, Wood, & Carberry, 2005; Hancock, Lesch, & Simmons, 2003; Owens, Wood, & Owens, 2007; Schlag, 1993; Shinar, Tractinsky, & Compton, 2005).

Other than these main effects – or alternative to them – interaction effects of age and task complexity are reported: with increasing environmental complexity and in comparison to younger drivers, older drivers disproportionately decreased speed, had increased reaction times or decreased driving performance (Chaparro et al., 2005; Hancock et al., 2003; Horberry, Anderson, Regan, Triggs, & Brown, 2006; Kramer, Cassavaugh, Horrey, Becic, & Mayhugh, 2007; Schlag, 1993; Shinar et al., 2005). Workload theories as introduced in chapter 2.3.5 can provide an explanation for these interaction effects: task demand only affects performance once a certain threshold is exceeded. For older drivers this threshold might be situated lower than for younger drivers due to the decrease in physiological and cognitive functioning (see above). However, because this age related decline shows a high inter-individual variation and cannot be equated to chronological age, a threshold for age cannot be named beyond which driving is unsafe (Ball, Owsley, Sloane, Roenker, & Bruni, 1993).

Often, driving experience is used as a single independent variable or at least as a correction factor when other taxonomic differences are analysed. Given that driving experience is seen as a continuum, the function describing the relationship between driving experience and accident involvement is annual mileage to the power of 0.25 (Maycock, 1997). Several differences can be found when experienced and inexperienced drivers are compared. These differences concern fixation patterns (Cohen, 1987), fixation duration (Chapman & Underwood, 1998), scan paths (Underwood, Chapman, & Brocklehurst, 2003) or steering wheel movements in curve negotiation (Cavallo, Brun-Del, Laya, & Neboit, 1988). In the latter study it was found that inexperienced drivers tend to use feed-back rather than anticipating feed-forward strategies. This finding is important insofar as it indicates that anticipating action is an important part of traffic safety (see also chapter 2.4.2).

In addition to demographic variables such as age and driving experience, so called traits constitute an important group of variables when it comes to modelling group specific driver and driving behaviour. Traits are behavioural dispositions with cross-situational consistency and stability over time and situation (Amelang, Bartussek, Stemmler, & Hagemann, 2006). Usually, traits are not assessed when accidents are investigated and so they are not integrated into accident data bases. Therefore, relating traits to driver and driving behaviour involves explicit testing of drivers, often with predefined hypotheses derived from the nature of the construct under investigation.

As a prototypical example of such traits, sensation seeking is introduced here in more detail because it is quite influential in the field of traffic (Herzberg & Schlag, 2003). Furthermore, it is not an isolated construct but is also correlated with several psycho-physiological and psycho-pharmacological parameters (Brocke, Strobel, & Müller, 2003; Jonah, 1997b; Zuckerman, 1994, 2007) and thus allows its interpretation in the context of other models. Sensation seeking is defined as 'seeking of varied, novel, complex and intense sensations and experiences and the willingness to take physical, social, legal and financial risk for the sake of such experience' (Zuckerman, 1994, 2007). Not surprisingly, sensation seeking is associated with risky driving (for an overview see Herzberg & Schlag, 2003; Jonah, 1997a, 1997b). However, as Zuckerman (2007) points out, this is not because high sensation seekers seek risk for risk's sake, but because they need novel and intense stimulation. This is supported by a study of Heino et al. (1996) who found that in general high sensation seekers rate situations as less dangerous than do low sensation seekers.

Also important in the context of driving and in relation to motivational models (see chapter 2.3) is the assumption that inter-individual differences are the result of differences in the tonic level of arousal. Such differences were also found for other differential variables (Eysenck, 1977). With respect to this tonic arousal level, Zuckerman (1976; 2007) assumes that high sensation seekers are under-aroused, whereas low sensation seekers are over-aroused. According to Zuckerman (1976; 2007) this results in a characteristic effect on both the inverse U-shaped performance-arousal function and the aversive, linearly increasing fear-arousal function. For low sensation seekers the aversive fear-arousal function is characterised by a steeper increase with increasing arousal. Due to their tonic over-arousal, further stimulation due to strong stimuli is perceived as aversive and is avoided. With respect to the effect on the performance-arousal function, the point of optimal performance is situated to the right on the arousal or risk abscissa for high sensation seekers.

2.3 Driving as a Self-Paced Task: Motivational Models

2.3.1 Introduction

In contrast to the longer-lasting differences between driver groups due to traits and demographic variables, motivational models emphasise transient states and stress the self-paced nature of the driving task (Ranney, 1994). A motive is a drive (Latin *movere*

= to move; to arouse) and motivation is defined as the entirety of motives in a given situation at a given time (Schlag, 2004). As every motive has both magnitude and direction, mathematically a motive can be seen as a vector (Lewin, 1982) and motivation is the sum of the different motive vectors. This resulting motivation vector has itself magnitude and direction and determines behaviour. Existing models differ, both concerning the number of motives which are used (one, two or many) and also in their approach concerning which of these are relevant to driving. The most influential group of motivational models so far are risk models.

In contrast to risk homeostasis with risk oscillating around a target risk (Wilde 1994, 2001), other authors see task difficulty or workload homeostasis as the most important motivational factor (Fuller, 2005; Hoyos, 1988; Hoyos & Kastner, 1987). Besides such single factor approaches, multiple factor models have also been proposed (Rothengatter, 1988). Based on the description of different conflicts by Lewin (1982), driving behaviour could also be seen as the outcome of a comparison between two (opposing) forces: an accelerating, pushing or pulling approach force and a decelerating, rejecting or slowing inhibition force (see also Zuckerman, 2007). A central aspect of motivational models – though not restricted to them – is the concept of behavioural adaptation.

2.3.2 Behavioural Adaptation

Behavioural adaptation describes the phenomenon in which people adapt their behaviour to changing situations or changing situational demand. The OECD (1990) defined behavioural adaptation as:

> (…) those behaviours which may occur following the introduction of changes to the road-vehicle-user system and which were not intended by the initiators of the change; behavioural adaptations occur as road users respond to changes in the road transport system, such that their personal needs are achieved as a result; they create a continuum of effects ranging from a positive increase in safety to a decrease in safety. (p. 23)

In this OECD report (1990) summaries of studies dealing with behavioural adaptation can also be found.

Whether the net outcome of road safety measures is positive or negative depends on the amount of unintended factors due to behavioural adaptation (Elvik & Vaa, 2004). Following risk homeostasis theory (see chapter 2.3.4), one could argue that behavioural adaptation implies that engineering measures alone would not result in a reduction of accidents. In fact there are publications supporting this assumption.

When comparing data over a 14-year period (1984-1997) from 50 US states it was found that the downward trend in fatalities was due to demographic factors, an increase in passive safety and improvements in medical technology (Noland, 2003). According to this publication, improvements in infrastructure sometimes even have negative effects. Infrastructure included total lane miles, average number of lanes, lane width and percentage of each road class. Curvature, shoulder width, separation of lanes and presence of roadside hazards were not included but it was implicitly assumed that newer roads are built in a safer way. Noland (2003) provoked with the conclusion: ‘Results strongly refute the hypothesis that infrastructure improvements have been effective at reducing total fatalities and injuries’ (p. 599).

Furthermore, Dulisse (1997) points out that the effects of behavioural adaptation are sometimes even underestimated due to methodological shortcomings such as the inclusion of drivers who wore seat belts even before wearing was made compulsory. However, Rothengatter (2002) states that adaptation in fact occurs but that the effects are not strong enough to negate positive impacts of safety measures. An example is the effect of road lighting where behavioural adaptation occurs but safety usually results in a net benefit (Assum, Bjørnskau, Fosser, & Sagberg, 1999).

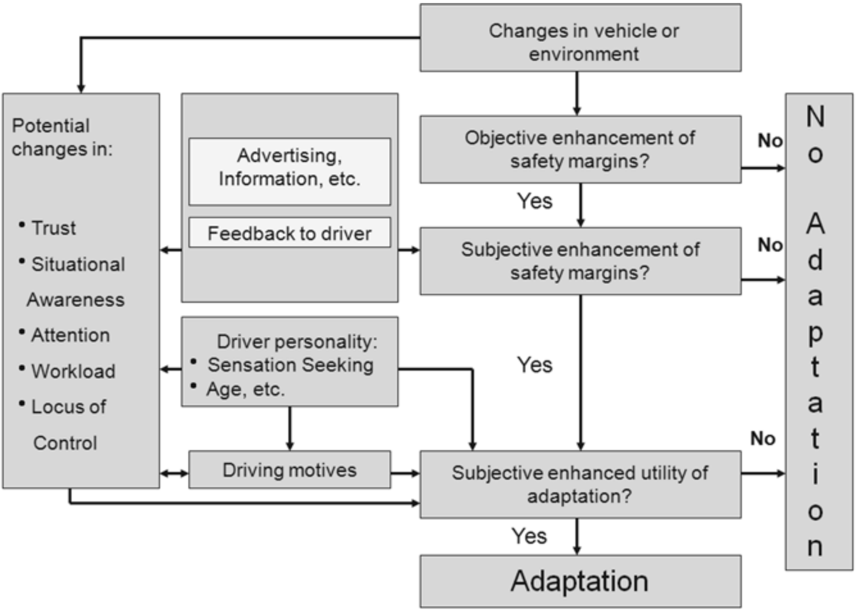


Figure 3. Process model of behavioural adaptation (Weller & Schlag, 2004).

The different findings concerning the amount of behavioural adaptation can be explained by the multiple factors that influence the occurrence of behavioural adaptation. These factors were summarised in a model developed by Weller & Schlag (2004) (see Figure 3). Similar aspects were named by Bjørnskau (1994; cited from Elvik & Vaal, 2004).

According to this model, a prerequisite for unintended behavioural adaptation to occur is that the changes in the vehicle or the environment result in an objective enhancement of the safety margins. This potential for behavioural adaptation must be perceived by the drivers and they must be aware of it. Whether this is the case depends on the communication of the measure through the media or advertising on the one hand and on direct feedback to the driver on the other hand. To result in adaptation, the change in behaviour must also be perceived as being beneficial to the driver (utility maximization). This utility maximization function is different between driver groups as well as within the same driver group. The latter depends, for example, on whether the drivers are in a hurry or not.

Independent of this chain of action (objective enhancement, subjective enhancement, utility maximization) is a second path which leads to adaptation. This second path is the result of changes in the nature of the driving task which affects several psychological variables. For example, speed might increase due to a decrease in task demand and workload, caused in turn by changes in the vehicle or the environment such as straighter or broader roads (see also chapter 2.3.9).

2.3.3 Control Theory Applied to Motivational Regulation

Some principles of behavioural adaptation can also be described more formally by using the definitions of control theory. Integrating a chapter on control theory seems appropriate as lateral and longitudinal control could indeed be seen as typical control tasks (Weir & Chao, 2007). Control theory describes the adjustment of an input signal in a dynamic system in order to achieve congruence between desired output and actual output.

Despite the fact it was originally developed for technical controllers, it was soon also applied to human controllers (overview in Jürgensohn, 2007). The German 'Handlungsregulationstheorie' (Action Regulation Theory) (Hacker, 2005) can be seen as a successful application of the principles of control theory (Hacker, 1994, 2003; Lord & Levy, 1994). Some authors even see (perceptual) control theory as a third significant theoretical framework to explain behaviour as an alternative to behaviouristic stimulus-response models on the one hand and cognitive models on the other hand (Taylor, 1999).

Control theories are applied in dynamic systems, either for goal directed human behaviour or for technical controllers which strive to achieve or maintain a predefined reference value or reference state. In its simplest form, in which the emphasis is on achieving not maintaining the reference value, open-loop control can be used (Figure 4):

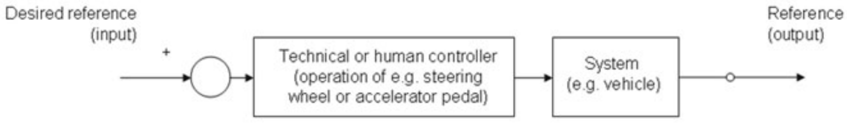


Figure 4. Schematic open-loop control system.

Open-loop control can be assumed in driving for an anticipatory adjustment of speed before entering a curve. If all curve characteristics were known exactly by the driver, then such open-loop control would be enough for driving. However, with the changing situation, the reference value might have to be adjusted to the new situation. Furthermore, the actual output value might deviate from the desired output value due to external disturbances or due to potential changes in the effectiveness of the control actions. To detect and subsequently diminish these discrepancies between predefined goal or reference value on the one hand and actual value of the system on the other hand, a feedback-loop is required (Hacker, 1994). The integration of a feedback-loop transforms the once open-loop control system depicted in Figure 4 into a closed-loop system depicted in Figure 5.

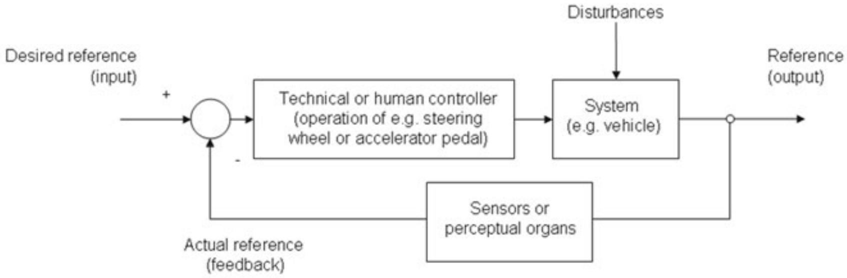


Figure 5. Schematic closed-loop control system.

Due to the multitude of external disturbances and the constant changing of a situation over time, driving is to a large extent indeed a closed-loop control task. In fact, it is almost impossible to drive in entirely open-loop control, simply because even in total occlusion conditions the driver uses ever present tactile feedback (Godthelp, Milgram & Blaauw, 1984). Nevertheless, there is evidence that both open- and closed-loop control are applied by human drivers even during simple lane change manoeuvres (Wallis, Chatziastros, Tresilian, & Tomasevic, 2007).

Such a two-process model was proposed and tested for steering in curve driving by Donges (1978). Open-loop control starts with an anticipatory steering action before entering the curve until the first maximum of steering wheel angle is reached. This open-loop control is guided by perceived curve characteristics and is situated on the manoeuvring or guidance level (see chapter 2.1). It was found that anticipatory steering action took place about 1.1 seconds before the vehicle passed the location of curvature change. According to Donges (1978) the curvature could best describe the driver's impression of the required steering action.

After anticipatory control, compensatory and corrective closed-loop control is applied. This corrective steering behaviour is quite automated and situated on the control level of the driving task. Visual information for this process is derived from three visual cues, namely lateral deviation, heading angle error and path curvature error. Path curvature error is derived from the velocity vectors in the optic flow field as described by Gibson (1986) (see chapter 2.4.4). Godthelp (1988) could further show that drivers use a constant time-to-line-crossing (TLC, see also chapter 2.4.4) irrespective of speed level for corrective steering actions.

For the regulation of speed, Reymond, Kemeny, & Droulez (2001) and Fuller (2005) similarly distinguish between anticipatory, open-loop control and reactive, closed-loop control. While Reymond et al. (2001) see lateral acceleration as a relevant input variable for the regulation of speed, the theory of Fuller (2005) is based on a homeostatic regulation of speed to assure a certain level of demand (see chapter 2.3.5). Such homeostatic regulation of behaviour is also assumed by Wilde (1988) who sees risk as the relevant variable (see chapter 2.3.4).

The latter examples show how control theory forms the underlying basis of motivational theories and can even be used in the context of behavioural adaptation (see previous chapter). How such motivational control can be used to detect unsafe speed regulation is explained through the example illustrated in Figure 6. In this prototypical example, workload is the relevant target variable. By adapting speed to the demand of the situation, the driver seeks to keep workload more or less constant at a certain level. In the case of curve driving, this requires an anticipated, open-loop reduction of speed well ahead of the curve (see left side in Figure 6). If such open-loop control fails, closed-loop control is required to abruptly reduce speed. This will considerably increase workload (right side in Figure 6).

While Fuller (2005) sees the ability for anticipatory control as an individual function of driving experience, the principle can also be transferred to road situations. ‘Good’ road design would allow the driver to correctly anticipate task demand. This would cause an anticipatory appropriate reduction of speed and would result in a more or less constant level of workload throughout the entire curve. Such application to the road situation has also important practical consequences as it allows the quality of road or curve designs to be tested. While mere differences in speed might be difficult to interpret, an increase in workload should be a non-ambiguous indicator of an unsuccessful adaptation of speed (see Figure 6). This in turn indicates bad design and could be regarded as critical to safety. Amongst others, this relationship was used as a basis for the formulation of hypotheses in later chapters.

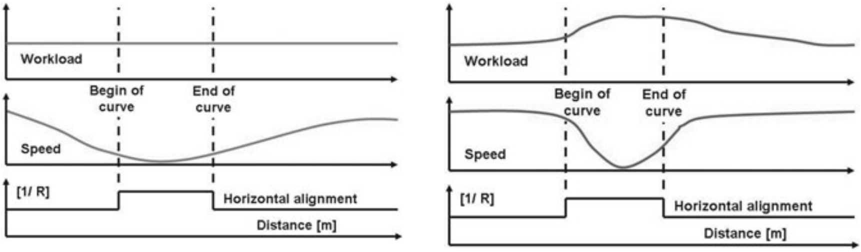


Figure 6. Hypothetical differences in speed and workload in curves with good (left) and inappropriate design (right) (modified from Fuller 2005).

However, the application of control theory in human behaviour is also criticised, mainly because it is regarded as too mechanistic (Locke, 1994). Furthermore, control-loop models which strive to exhaustively define driving require a large amount of variables (see the control loop models of Dilling, 1973; Durth, 1974). Because it is usually unknown how these variables interact with one another, such models can merely be descriptive in nature and cannot be validated statistically. However, by replacing behavioural variables such as speed with motivational variables such as subjective risk or workload, as shown above, these shortcomings might be partly overcome.

2.3.4 *Risk Models*

A basic yet important distinction should be made between subjective and objective risk when discussing risk models. Klebelsberg (1982) defines objective risk as the measurable probability of having an accident, whereas subjective risk is the risk of having an accident estimated by the driver through the perception of the road environment. According to Klebelsberg, situations are unsafe as soon as subjective risk is lower than objective risk. This is because drivers adjust their behaviour according to subjective, not objective risk.

The concept of subjective risk as a relevant mechanism for driving behaviour was further developed by Wilde. Originally called risk homeostasis theory (RHT) (Wilde, 1988) it was later termed target risk theory (Wilde, 1994, 2001). In short, the theory states that accident-rates per unit of time remain equal despite objective improvements in the driving environment, both concerning infrastructure and vehicles. The assumed reason is that drivers adjust their behaviour so that their subjective risk equals a more or less constant target risk (see Figure 7).

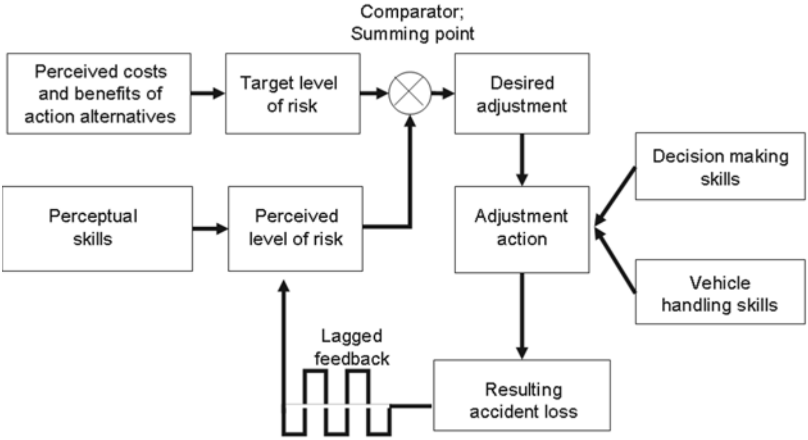


Figure 7. Risk homeostasis theory according to Wilde (1994).

According to Wilde (1994; 2001) the height of the target risk is defined by combining the two utility functions for both comparatively risky and comparatively safe behaviour:

- the expected advantages and the expected costs of comparatively risky behaviour, and
- the expected advantages and the expected costs of comparatively safe behaviour.

Weighing up these utility functions is not done consciously but rather seen as a highly automated unconscious process (Wilde, 1994, 2001).

While Wilde's ideas are highly influential in traffic psychology, they usually provoke equally widespread criticism at the same time. Before summarizing the criticism, an attempt is made here to understand the empirical background which led to the development of RHT in the first place. This is done in quite some detail because it also serves as an illustration of different issues with relation to the empirical part of this thesis.

A central study for the development and understanding of the concept of a target level of risk as used by Wilde is a study conducted by Taylor (1964). This study is cited in a number of influential motivational driving behaviour theories but interpreted differently by each author (Fuller, 2005; Groeger, 2000; Näätänen & Summala, 1976; Wilde, 2001). Therefore, it seems appropriate to summarise the central findings of this study together with a short discussion of the implications. The necessary discussion of electrodermal activity (EDA) and its parameters can also be used as an exemplary illustration of the possibilities and limitations of using psychophysiological measures in risk (or workload) assessment in driving.

In the article (Taylor, 1964), Taylor reports two real road driving studies in a predominantly suburban environment with a variety of road conditions. Participant numbers were twelve for the first and eight for the second experiment. Selected sections from the total driven courses (2 × 37 miles and 21 miles) were used for the analysis. The lengths of these selected sections were 12 and 16 miles which were further subdivided into sections of different length but similar characteristics (40 sub-sections in the first and 19 sub-sections in the second experiment). For these sub-sections an aggregated value of each dependent variable was calculated. The dependent variables were different parameters of the galvanic skin response (GSR) taken from the fingers and speed. Speed was calculated from the distance driven and the time needed, whereby time needed was measured via a stopwatch. The independent variable in the study was the accident-rate of personal-injury accidents (definition of accident rate and other accident measures in FGSV, 2003). Further independent variables were the number of side-turnings per mile of road[1], traffic flow (traffic density) and the number of lanes. To identify the influence of the independent on the dependent variables, regression analyses and analyses of variance were used in combination with correlation analyses.

Before reporting the results, the electrodermal parameters used for the analyses need some explanation. First of all, the term GSR is today usually replaced by the generic term electrodermal activity (EDA) which describes the phenomenon more appropriately (Schandry, 1998). In general, tonic and phasic changes of EDA are

---

[1] Taylor once also uses the term 'junction' instead of 'side-turnings'. Whether the driver actually turned at these side turnings or whether these are just geometric elements without the driver actually turning there is not clear. Boucsein (1988) used the term 'Abbiegevorgang' (indicating that drivers actually turned off) when summing up Taylor's study.

differentiated. To assess tonic changes, the skin conductance level (SCL) and the number of non-specific fluctuations or non-specific electrodermal – or skin conductance – responses (NS.EDRs or NS.SCRs) are used. Phasic changes are measured by recording skin conductance responses to specific stimuli (EDRs or SCRs).

These two kinds of EDRs are distinguished depending on whether an identifiable stimulus is present or not. Phasic changes occur with a latency of 1 to 3 seconds after stimulus presentation (Boucsein, 2001; Dawson, Schell, & Filion, 2007). Non-specific or non-stimulus bound fluctuations (NS.EDR) occur without an external stimulus and are seen as indication of general activation (Boucsein, 2001). As both responses cannot be distinguished from their form in the recorded data, the analysis and documentation of potential stimuli is needed. Therefore, the first step towards correctly interpreting the results of Taylor in the context of RHT is to identify which of the two parameters were used in the experiments. Segregating both parameters is seen as important because stimulus-induced EDR cannot necessarily be attributed to subjective risk or risk related arousal, but might also have other causes not relevant to driving (Näätänen & Summala, 1974).

With relation to the EDR/GSR, Taylor writes that 'It was usually possible to observe external events which could have caused the responses …' (Taylor, 1964, p. 442). Given this quotation, EDR could only be indicative of subjective risk, if these 'external events' were somehow related to risk. In this case these stimuli would serve as cues which result in an EDR. On the other hand, Taylor reports that ' … the general frequency of responding … tended not to vary with road conditions' (p. 442) which is further clarified with 'For instance, if responses were occurring irregularly at an average frequency of, say, three in five minutes, this would continue no matter whether the driver was in heavy traffic or on open roads' (Taylor, 1964, p. 442). The last quotation supports the use of EDRs as NS.EDRs which is in line with the summary of the study in Boucsein (1988).

In my opinion it cannot be determined from Taylor's article (1964) whether stimulus dependent or stimulus independent EDR were recorded. A second important aspect with relation to the first concerns the parameter used to quantify EDRs. In principle, several parameters can be used to do so. However, the difficulty to clearly define the start and end of an EDR resulted in the amplitude of the EDR being the most common parameter (Boucsein, 2001).

While NS.EDR can be described by the same parameter(s), NS.EDRs are usually aggregated and their number within a certain time (their frequency of occurrence) is reported. Taylor first calculated the integral of the EDRs, which can be seen as a combination of both amplitude and development of the single EDR over time. The 'increment in the GSR integral' was divided either by the time it took to negotiate a mile of the section or by the distance of the section. The resulting parameters were called:

- GSR rate (when divided by time), and
- GSR per mile.

GSR per mile depends on the speed driven. Provided that the GSR rate is not affected by changes in speed per se, then higher speed should result in a lower value for the parameter GSR per mile.

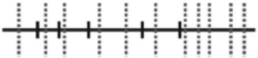
The argument in favour of RHT is based on a combination of results relating to GSR per mile and GSR rate. As GSR rate is implicitly equated with target risk, it is somewhat surprising that mainly results concerning GSR per mile are reported in Wilde's books (Wilde, 1994, 2001). Results concerning GSR rate cited in favour of RHT are mainly deduced from the GSR per mile results (summarised in Table 1) or are based on non-significant differences between the sections. The matrix in Table 1 shows the significant correlation coefficients of the 40 section-wise values of GSR per mile (i.e. the spatial distribution of GSR) and speed, both averaged across participants and sections, and the accident-rate and side turnings, both per mile and also averaged across sections.

Table 1. Significant correlation coefficients between independent variables and GSR per mile (Taylor 1964).

	GSR per mile	Accident-rate per mile	Number of side turnings per mile
Accident-rate per mile	**+ 0.61**		
Number of side turnings per mile	+ 0.67	+ 0.68	
Participants average speed	**- 0.75**	**- 0.67**	- 0.80

The strong support Wilde deduces from this correlation matrix in favour of RHT is based on the significant mutual correlations and their polarities between accident-rate, speed and GSR per mile (printed in bold). According to Wilde, the negative correlation between GSR per mile and speed means that drivers kept the GSR rate (GSR per unit time) more or less stable. This interpretation is supported by Taylor (1964) who reports that 'no systematic variation' (p. 445) could be found between GSR rate and average speed as the result of a regression analysis. According to Taylor, the reason is that 'a constant temporal rate of activity will be sparsely distributed over the terrain if the driver's speed is high, and densely if it is low' (Taylor, 1964, p. 445) (for an illustration see Table 2).

Table 2. Idealised prototypical example based on Taylor's results and Wilde's interpretation for illustrative purposes.

Idealised bird's eye view of a road of 1 km length with side turnings (vertical black bars) and GSR (dotted red vertical bars)	Side turnings	Accident-rate	Speed [km/h] ([m/s])	Execution time in s	GSR per mile [1/km]	GSR rate [1/s]
	5	8	50 (14)	70	10	0.14
	2	2	100 (28)	35	5	0.14

Referring back to the discussion of EDA above and the preceding interpretation by Taylor, EDRs which are the basis of what Taylor calls GSR, would represent non-stimulus arousal and thus are somewhat inconsistent with the fact that external events were found to be responsible for them (see above). However, what Taylor and Wilde assume is that the distribution of these external events (and accidents) is also evenly distributed over time once speed is accounted for. This would result in 'side turnings per minute' or 'accidents per unit time of travel' (Taylor, 1964, p. 447). In his book, Wilde does not discuss the fourth variable in Table 1 (side turnings per mile) and the polarities of the correlation coefficients with the other variables.

Although no causal relationship can be inferred from a mere correlation, the signs in Table 1 can be summarised as: the more side turnings, the more accidents, the higher GSR per mile and the lower speeds. If these results were interpreted in favour of RHT it has to be assumed that the correlations would diminish, once corrected for speed and referenced to time rather than distance. Is this idea supported by the results presented by Taylor concerning GSR rate?

GSR rate was averaged between 'physically similar sections' resulting in six section group values. No significant differences in GSR rate were found between these sections, whereas significant differences were found between the eight participants (second experiment, see above). This is in line with the argument of RHT: individual differences in 'target risk' but constant risk between different road sections. On the other hand, Taylor reports that there were indeed some systematic differences when non-aggregated section values were compared. For example, higher GSR rates (not GSR per mile) were found in short sections with intersections, and in sections with roundabouts (the latter effect depending on whether other traffic had to be crossed or not).

Furthermore, not only were GSR rates higher, but also the accident-rates at these sections. As GSR rate (indicative of 'target risk') should be more or less constant, this finding constitutes an inconsistency in Wilde's argument. On the other hand, in line with Wilde's theory of a target risk, it could be argued that the accidents are higher precisely because GSR rate is higher, indicating that the usual mechanisms to keep target risk constant did not apply at junctions. This argument is shared by Taylor who states that junctions are 'discontinuities in the hazard situation' (Taylor 1964, p. 449) which might be passed too fast to be noticed by the GSR risk monitor. However, if target risk does not 'work properly' in a systematic way at certain locations (indicated by the higher GSR rate and higher accident-rate at junctions) then why is it needed at all? In fact, the accumulation of accidents at a certain location could also be easily explained by Klebelsberg's distinction between objective and subjective risk (see above).

When it comes to evaluating the overall quality of the results per se and their interpretation within RHT, Taylor reports that 'the considerable variation in GSR rate observed … , cannot yet be interpreted as true differences in the level of anxiety of the subjects. They reflect, perhaps, shortcomings of the GSR technique which will require further investigation' (Taylor, 1964, p. 449). An indication of possible measurement artefacts are reported by Taylor himself when stating that GSR rate increased with elapsed time in each individual drive. While this increase was at the same time independent of a general increase in skin conductance (likely to be SCL, see above), it indicates a general problem of using EDA parameters while driving.

EDA parameters vary with temperature, respiration (especially deep breaths) and most importantly with movement (Boucsein, 1992, 2001; Schandry, 1998). Movement especially is seen as a highly critical influence when EDA is recorded in real driving, especially when movement is not controlled. Such control can be achieved for example by forcing the participant to drive with one hand only or by deriving the values from the passive foot when driving a car with automatic transmission (as is reported in Richter, Wagner, Heger, & Weise, 1998; Wagner, 2000). No such control is reported by Taylor, nor is it reported how the artefacts were controlled and how exactly GSRs were identified and subsequently 'electromechanically' integrated. In my opinion, the results concerning the correlation between junctions and GSR per mile and GSR rate (see above) could be such a movement (i.e. steering) artefact in the case of drivers actually turning off the road into another road (see footnote above).

Last but not least, the validity of EDA parameters, especially EDRs, to measure subjective risk as well as their discriminative validity with respect to other psychological constructs such as attention, emotion and mental workload needs some consideration. According to Boucsein (1995) EDA parameters are indeed related to the intensity of negative, fear-related arousal which could be interpreted as subjective risk. On the other hand, it is agreed upon that EDA is indicative of a variety of psycho-physiological processes which is mirrored in the large variety of studies in which it is used (Boucsein, 1995, 2001; Collet, Petit, Priez, & Dittmar, 2005; Dawson et al., 2007; Hancock & Verwey, 1997; Verwey & Veltman, 1996). Accordingly, EDA as applied by Taylor (1964) can be used as an indicator of subjective risk, whereas other authors interpret it as indicator of workload (Groeger, 2000). But even when applied to the measurement of workload, the diagnosticity of EDA measures to distinguish between different aspects of workload (ISO 10075-3, 2004; Wierwille & Eggemeier, 1993) is regarded as low (de Waard, 1996; Wagner, 2000).

Taken together it can be stated that despite its valuable and innovative approach the study by Taylor (1964) has some methodological shortcomings. With the central role of this study in the framework of target risk and target risk 2 (published almost 40 years after Taylor's study), it must be concluded that RHT possibly cannot be regarded as sufficient to constitute a driving theory for rural roads.

Independent of the study by Taylor (1964), RHT provoked much wider negative reactions when it was published. These reactions were and are a result of the implications of the theory with respect to traffic safety. The idea that drivers adjust their actions to keep a target level of risk means according to Wilde, that

- merely improving the objective safety of cars or the infrastructure will in the long run not result in a decline in accident figures when referenced to units time travelled, and
- the only way to improve safety is to lower the target level of risk.

Not surprisingly the theory is controversially debated and criticised for some of its assumptions (Elvik & Vaa, 2004; Evans, 1986; McKenna, 1988). A major criticism of RHT is that the theory supposes that (individual) target risk can be measured in terms of characteristic accident figures. According to Vaa (2007) this is unlikely and was never proven. Furthermore, it is difficult to distinguish between the terms used which is reflected by the fact that target risk, feeling of risk and effort (see next chapter) are sometimes used interchangeably in one and the same study (Heino et al., 1996). It should also be noted that in the study by Heino et al. (1996), feeling of risk is used instead of estimated accident risk.

Evans (2004) in particular stated that RHT is already refuted by accident data. However, upon further investigation, Evans' argument is not entirely convincing as the accident data presented by him (Evans, 2004) in support of this argument is not only sparse but also does not fully take into account the influence of speed and thus the time related conversion of accident-rates on which the argumentation of Wilde is based.

Despite the criticism of Wilde's theory, most researchers agree that it has had a positive impact for understanding driving behaviour and that it has revealed important mechanisms which can be used to explain accidents (Elvik & Vaa, 2004; Janssen & Tenkink, 1988).

Another concept of risk in driving – this time on the individual level – was proposed by Näätänen & Summala (1974; 1976). In contrast to Wilde, Näätänen & Summala claim that subjective risk during driving is usually nil. Similar to Wilde, subjective risk is determined by the subjective probability and the subjective significance of an adverse event (SEU). For Näätänen & Summala, an adverse event could be either an accident or being caught by the police. In contrast to Wilde's RHT, in which subjective risk is an excitatory, motivating force (see the term 'target' risk), it is an inhibiting force in the zero risk model of Näätänen & Summala. Until the threshold of zero subjective risk is exceeded, risk does not play a major role in driving. Rather it is monitored in the background by the so-called 'subjective risk monitor'. Only when subjective risk exceeds zero does this subjective risk monitor signal to the driver to change his/her behaviour. In this case, Näätänen & Summala termed the decision for a certain driving behaviour or manoeuvre 'reactive', whereas it is called 'active' when the driver can choose the desired behaviour without interference from the risk monitor.

For the evaluation of the model it is important to note that the authors suppose that the risk monitor supervises both the subjective risk in the present situation as well as the expected subjective risk for the situation ahead. Instead of a certain target level of subjective risk, the excitatory force behind driving is seen in 'motivation' (Näätänen & Summala, 1974, 1976). Motivation can be seen as the entirety of motives at a given time (see also beginning of the chapter). Näätänen & Summala (1974; 1976) cite studies which show the influence of motivation on perception, expectation and the amount of perceived subjective risk associated with a particular behaviour in a given situation.

This does not mean that drivers are seen as being motivated to seek risk for risk's sake but that they might underestimate the probability or severity of an accident because of a strong motive, for example, to drive fast (expressed by 'being in a hurry'). The authors did not find much literature to support this idea but the later research on sensation seeking could be cited as evidence in favour: rated risk depends on the height of sensation seeking (Heino et al., 1996) (see chapter 2.2).

Besides motivational factors, perceptual shortcomings and a general overestimation of one's abilities lead to an underestimation of risk, which is seen as the cause behind accident occurrence. Again, the inferred postulation for road safety would be to increase subjective risk, at best by simultaneously increasing objective safety (without the driver noticing it).

Although to all appearances risk theories are understandable, Fuller (2005) points out that the sole distinction between objective and subjective risk cannot explain driver behaviour. The reason is that according to Fuller subjective risk as used by Klebelsberg (1982), Wilde (1988; 1994; 2001) and Näätänen & Summala (1974; 1976) is the outcome of a conscious cognitive process to estimate objective risk, which is different from the feeling of risk.

Fuller (2005) defines feeling of risk as an emotional response to threat. In contrast to subjective risk of an accident which is not supposed to change until a certain threshold is reached, feeling of risk is seen as a continuum. Due to this difference, only the latter could explain behaviour below the threshold level for subjective risk. However, given this criticism, simply replacing ‘subjective risk’ by ‘subjective feeling of risk’ might in itself weaken this kind of criticism. Nevertheless, Fuller (2005) developed his own theory, based on demand and workload.

2.3.5 Workload Models

The multitude of different road characteristics, the various features of the landscape through which these roads lead and the diversity and number of other road users and environmental conditions mean that the characteristics of the driving task change constantly. These characteristics result in a certain level of physical and also mental demand or stress which impacts the driver when negotiating this road. Mental in this sense includes all cognitive, informational and emotional processes of human experience and behaviour. Mental stress or demand is defined as ‘the total of all assessable influences impinging upon a human being from external sources and affecting it mentally’ (ISO 10075-1, 1991, p.1). Acting within a stressful environment or executing a demanding task will have an effect on the person who executes this task. This effect will vary with the characteristics of the task and within an individual, depending on its state, and between individuals, depending on trait or demographic differences (see chapter 2.2).

In general, short- and long-term effects are distinguished. Long-term effects of inappropriate demand can result in illness (Richter & Hacker, 1998). As long-term effects are not seen as being particularly problematic for safe driving on rural roads, it is not a subject of this thesis. Mental workload denotes the short-term effects of demand and is consequently defined as ‘the immediate effect of mental stress within the individual (not the long-term effect) depending on his/her individual habitual and actual precondition, including individual coping strategies’ (ISO 10075-1, 1991, p.1).

Depending on task demand, workload and human performance interact in a characteristic way. De Waard (1996) has summarised the findings of different authors regarding this interaction. He describes the relationship between task demand and workload as U-shaped function, and the relationship between demand and performance as inverted U-shaped function (see Figure 8).

In driving, demand is foremost a function of the objective road and environment characteristics. Following the relationship depicted in Figure 8, best performance can be assured by designing roads and their environment in a way which corresponds to medium demand. In order to do this, the processes which lead to the relationships depicted in Figure 8 must be understood.

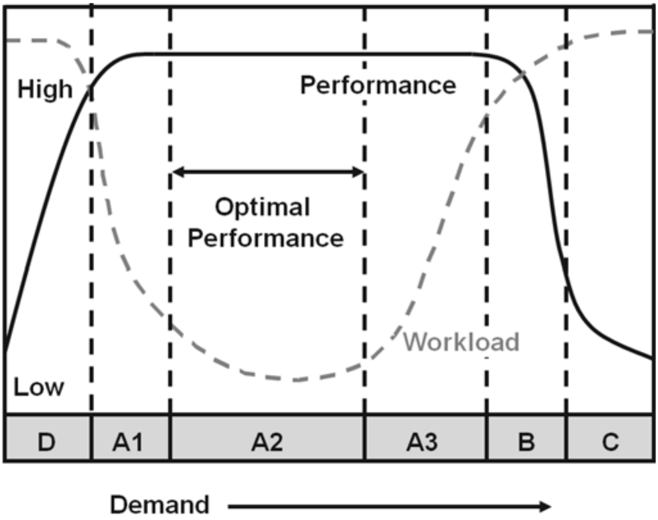


Figure 8. Interrelations between workload and performance on different levels of demand as developed by de Waard (1996).

Theories and concepts useful in this context exhibit a very close relationship to the construct 'attention'. This is because the same experimental paradigm (dual-task paradigm) is often used to describe effects of both workload and attention on performance or more precisely either workload or attention depending on the focus of the author. The resulting difficulties to distinguish thematically and structurally between workload and attention in a clear-cut way led to the following solution for the thesis at hand: concepts unique to attention are described in the chapter on attention (chapter 2.4.3), while concepts pertaining to both attention and workload are described here.

Because the relationships depicted in Figure 8 are central to the understanding of the effect of workload for safety they are explained here in more detail. In comparison to region A, region B shows a gradual decline in performance with increasing demand until a minimum in performance and a maximum in workload are reached in region

C. Theories which explain this phenomenon are typically resource theories (overview in Sanders, 1997).

Resource theories assume that resources are needed for task execution, that more resources are needed with increasing demand and, most importantly, that these resources are limited. Regarding the terms used, the upper limit of resources can be called capacity (for a discussion of the terms see de Waard, 1996, for an application of the term 'capacity' see Fuller, 2005, below).

Despite being generally agreed upon that there is in fact an upper limit beyond which no increase in capacity is possible, capacity does change depending on task demand within a certain range. This notion was proposed quite early on by Kahneman (1973) and was elaborated for under-load conditions by Young & Stanton (2002, see below). Workload is equated to the inverse of resources needed for executing the task. Resource theories assume either a single general capacity (Kahneman, 1973) or several independent resources (Wickens, 1984, 1991, 2008).

In his cube-like model, Wickens distinguished between resources according to the presentation of the stimulus (modality), the kind of task presented (codes), the cognitive stage involved, and the kind of response to the stimulus or task. In relation to driving, the most important assumption is that the different resources are independent of each other. The less resources two tasks have in common, the less interference there will be and the more capacity is left for each single task.

The phenomena in the regions B an C in relation to region A can also be explained with a second class of theories named bottleneck theories (overviews in Luck & Vecera, 2002; Meyer & Kieras, 1997; Navon & Miller, 2002; Pashler, 1994). These theories are usually used in relation to attention (see chapter 2.4.3) but also explain the degradation in performance as shown in region B and C. In this case, an increase in demand could be equated to an increase of information.

For example, such would be the case if speed was increased. Bottleneck theories assume a 'stubborn' (Pashler, 1994) bottleneck in information processing. Thus, an increase of information (see above) beyond bottleneck capacity will result in an increase of workload (provided that the driver is aware of the oversupply of information) and a subsequent decrease in performance.

Where in the different stages and processes of information processing this bottleneck is possibly situated is the subject of ongoing debates. Broadbent (1958) assumed that information from the environment is filtered at an early stage even before the information is processed, whereas Pashler (1994) assumes a single bottleneck at response selection. Other theories assuming late selection are discussed in Pashler (1998). In addition to the location of the bottleneck, and similar to the distinction between single and multiple resource theories, it is debated whether there is one (Broadbent, 1958) or many channels (Allport, Antonis, & Reynolds, 1972) that result in a respective number of bottlenecks.

The research paradigm used for bottleneck theories is the psychological refractory period paradigm (PRP-paradigm). In this paradigm, two stimuli which require a response are presented consecutively. The period between presentation of stimulus one and stimulus two is called stimulus onset asynchrony (SOA). The execution of such tasks results in a characteristic delay of the reaction to the second stimulus, which depends on the length of the SOA. An application of the SOA or PRP paradigm in (simulator) driving was recently published by Levy, Pashler & Boer (2006). The workload function shown in Figure 8 for the regions B and C in comparison to region A would have to be inferred from the degradation in task performance. If collected as a subjective rating, workload would have to be perceived by the individual either via the degraded task performance (as deviation from the desired task performance) or as a result of the 'effort' (see below) which has to be invested to ensure that the relevant information passes the bottleneck.

So far, two theories have been introduced which can be used to explain the effects of task demand on performance and workload as shown on the right side of Figure 8. However, none of these theories offer an explanation why performance should decrease with decreasing demand as depicted on the left side of Figure 8. Energetic theories are capable of doing so. These theories date back to the beginning of experimental psychology and even the inverse U-shaped function of sensation (pleasantness) and stimulus strength described by Wundt (1874) could be mentioned here. However, usually Yerkes & Dodson (1908) are cited as describing the relationship between arousal and performance as an inverse U-shaped function[2]. A similar relationship between human arousal and performance is proposed by Hebb (1955)[3].

The left side of the inverted U-shaped function in Figure 8 describes the decrease in performance and the increase of workload in region D ('deactivation', de Waard,

---

[2] Yerkes and Dodson (1908) investigated the strength of a stimulus needed to provoke a certain response. The experimental setup to investigate this was dancing mice which were required to select a white box from a white or black box. When they chose the black box an electric shock was administered to the mice. The current was varied in its strength (low, medium, high) in different experimental sessions and it was measured how long it took for the mice to learn which box to choose. The experiment was varied further by changing the light conditions of the two boxes. These changes made the two conditions more or less discriminative and thus the task as such more or less difficult (three experiments: medium, easy, difficult to detect the differences). The different experimental runs resulted in different arousal performance curves, suggesting that performance was best at medium levels of arousal and that with difficult conditions, this optimum level of arousal was reached at lower levels of arousal than when the task was of medium difficulty.

[3] Hebb used the term 'cue function' instead of performance on the ordinate. Cue function can be interpreted as the effectiveness of a certain stimulus to cause certain behaviour and is thus often equated to performance.

1996). As was already indicated when introducing resource theories, resources seem to decrease with decreasing task demand. Young & Stanton (2002) showed that the amount of capacity available for task execution diminishes when task demands are too low. They tested their hypothesis by calculating the ratio of attention to a visual-spatial secondary task. This ratio was measured by the gaze duration at the secondary task and the number of correct answers to this task. By analyzing this ratio they found that the allocation of attention to the secondary task becomes less efficient when demand decreases. This condition is often termed under-load. Under-load is assumed to be even more critical than overload, simply because it is much more difficult to detect (Young & Stanton, 2002).

It should be noted that the simple bell-shaped arousal performance function has evoked some criticism (Neiss, 1988) (with a reply from Anderson, 1990). Furthermore, today it is known that arousal is just one energetic mechanism which determines performance and workload. Pribram & McGuinness (1975) and Sanders (1983) distinguish between three systems, arousal, activation and effort. Arousal is related to the receptivity of sensory input while activation is seen as preparedness for motor-responses. Arousal and activation are not always distinguished in this way: either the terms are in general used interchangeably as for example acknowledged by Barry, Clarke & McCarthy (2005), or several (more than two) separate energetic mechanisms are distinguished which are all subsumed under the term arousal (Robbins, 1997).

Regardless of the point of view, all authors agree that at the very least a tonic and a phasic (stimulus dependent) component of arousal must be distinguished. VaezMousavi, Barry & Rushby (2007) named these components baseline and activated arousal which makes the distinction intelligible at first sight. Tonic arousal is not only assumed to vary with state (asleep/awake) but also to characterise trait-specific differences between individuals (Eysenck, 1977; Zuckerman, 2007). In relation to mental workload, it is important that activated arousal (as defined here) depends not only on baseline arousal and the stimulus characteristics (bottom-up) but also on a top-down component. This top-down component is usually subsumed under the term effort or effort system (Brocke, Tasche, & Beauducel, 1997; Pribram & McGuinness, 1975; Sanders, 1983).

During conscious processing and evaluation of information, and for compensatory control, effort is invested to adjust the level of arousal or performance[4]. The amount of effort needed to achieve a certain level of arousal or performance determines workload. Referring to Figure 8 this is the case in sections A1 and A3. Whereas in section A1, increasing effort has to be invested with decreasing demand, in

---

[4] The distinction between arousal and performance in this context is more a question of the researcher's focus: for physiologists, arousal is the important parameter whereas it is performance for cognitive oriented psychologists.

region A3 increasing effort has to be invested with increasing demand. Because insufficient arousal is the underlying challenge for performance in region A1, the effort exerted there is termed state-related effort (Mulder, 1986 cited in de Waard 1996). Similarly, the effort exerted in region A3 is termed task-related effort as effort is needed here to compensate the increasing demand of the task (Mulder, 1986 cited in de Waard 1996).

State-related effort can be measured with physiological parameters which indicate a general increase in arousal. In contrast, physiological parameters of task-related effort should be associated with the specific resources required by the task under investigation. The fact that effort and arousal are associated with different physiological systems offers another possibility to distinguish those components. While arousal is usually assessed with autonomous nervous system parameters such as cardiac or electrodermal activity (Barry et al., 2005; VaezMousavi et al., 2007), changes of the central nervous system as reflected by EEG parameters are used to assess activation of the effort system (Beauducel, Brocke, & Leue, 2006; Brocke et al., 1997; Fischer, Langner, Birbaumer, & Brocke, 2008).

One could argue that an increase in workload is of no relevance as long as there is no decrease in performance. However, this is not the case as an increase in workload, such as in regions A1 and A3, diminishes 'spare capacity' and thus by necessity limits the amount of resources available to compensate a sudden increase in demand. In the situation where demand exceeds spare capacity, a sudden, massive degradation of performance as indicated by driving errors or even accidents results. Furthermore, increased workload leads to negative effects when maintained for a longer period of time as in vigilance tasks. Several empirical studies involving driving are presented by de Waard in support of this model (de Waard, 1996, 2002).

(Mental) demand was defined above as all external sources impinging upon the human. One of these sources is the task itself. While often the demand associated with a task is a task-inherent characteristic, the case is somewhat different for driving. Driving is a time-critical task (see TTC & TLC, chapter 2.4.4) and thus its demand changes with speed. Because speed is usually under driver control, the driver is able to control the demand of the driving task, at least to a certain extent. The self-paced nature of the driving task allows motivational processes to take effect.

Summala (1997) proposes a hierarchical model of behavioural adaptation (see chapter 2.3.2), which is in fact a workload model where time margins constitute the central variable (see Figure 9). It is hierarchical in so far as it does not merely concentrate on lower levels of the driving task but also includes the navigational level. According to the model, the causal factor for trip and speed decisions is the available time at all levels of the driving task (see chapter 2.1). At the tactical and operational level, time margins are synonyms for safety margins and influence the action taken and the speed chosen. In case of short time margins, workload increases which leads to a modification (here: lowering) of the target speed. On the other hand, high time margins per se are not necessarily safe. In the situation where the driver feels to be in an under-load condition and cannot increase speed to decrease time margins, the driver might engage in secondary tasks. This in turn will ultimately lead to overload. In the situation where the driver does not engage in secondary tasks, the risky under-load condition will be prolonged.

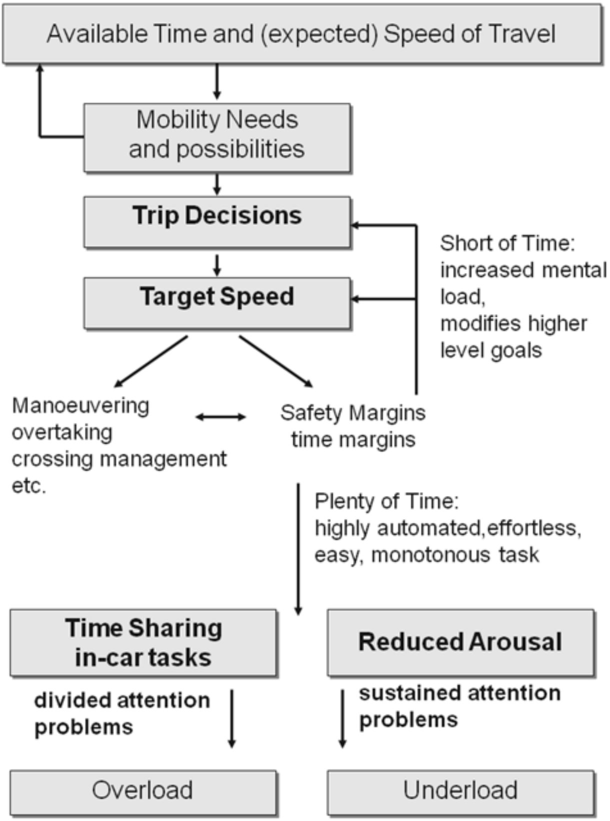


Figure 9. Hierarchical model of behavioural adaptation as proposed by Summala 1997.

Although accidents are not explicitly mentioned in the model, one could assume that both over- and under-load will result in accidents. Cnossen (2000) and Cnossen et al. (2004; 2000) investigated some assumptions of the theory and found that secondary task performance did not improve in less demanding situations although effort (or workload) indeed depended on speed.

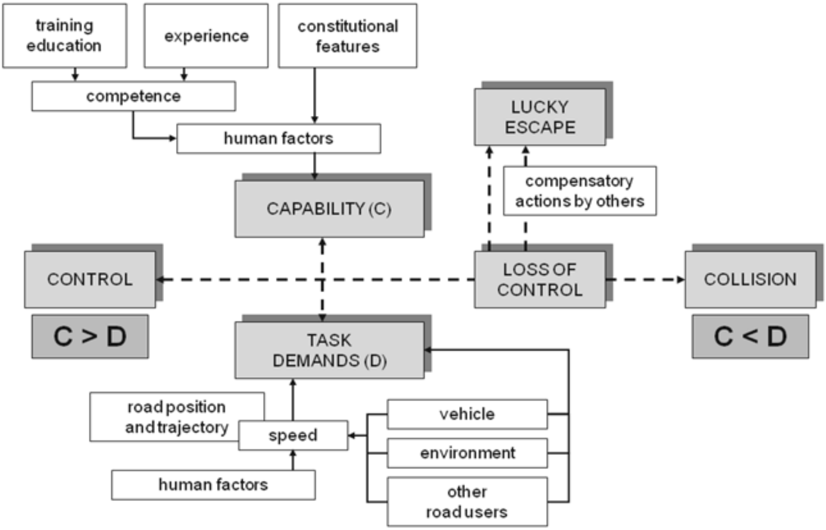


Figure 10. Task-capability interface model modified from Fuller (2005).

In his task-capability interface model (TCI-model, Fuller, 2005), Fuller directly linked the consequences of a mismatch between task demand and driver capability to safety (Figure 10). The steady interaction between task demands and capability (similar to available resources) results in safe driving or control as long as capability exceeds demand. Both task demand and capability determine task difficulty, whereby 'task difficulty is inversely proportional to the difference between task demand and driver capability' (Fuller, 2005, p. 463). Task difficulty in turn 'may also be considered to be equivalent to mental workload' with an additional physical workload component (Fuller, 2007, p. 174).

This model is essentially influenced by Wilde's RHT (see chapter 2.3.4) in that it similarly assumes a homeostatic process. In contrast to Wilde, Fuller (2005) claims that it is not risk but rather task difficulty which is the relevant variable behind this homeostatic process. By regulating speed, the driver has a powerful means to adjust task demand in a homeostatic way in order to maintain a certain (target) level of task difficulty.

The idea of workload homeostasis was also expressed by Hoyos & Kastner (1987) and Hoyos (1988) who, by referring to Lazarus (1991), assumed and tested (see also chapter 2.3.9.3) that drivers keep a 'dynamic' equilibrium between stress and strain. A similar approach was explained in Gstalter & Fastenmeier (1995).

Zeitlin (1998) also claimed to have found indications of workload homeostatic processes in driving. The empirical basis for this claim was a comparison of drives on real roads of different categories (rural, expressway, urban and city). However, while the statistical evidence of differences between these categories is overwhelming (results are given for speed, brake actuations per minute and two secondary tasks) the evidence for a homeostatic regulation of task difficulty is only given anecdotally.

An indirect evaluation of workload homeostasis is reported in a study by Jamson & Merat (2005). Jamson & Merat (2005) tested the effect of simulated in-vehicle information systems (IVIS) on driver and driving behaviour on rural roads. They found evidence '…that drivers, either consciously or subconsciously, developed a strategy to reduce primary task load whilst performing concurrent secondary tasks. This was shown by a significant reduction in driving speed during interaction with both the auditory and visual tasks' (p. 93). On the other hand, 'the success of this strategy is questionable' (p. 93) as can be seen when TTC to a breaking lead vehicle decreased despite decreased speed.

What evidence is given by Fuller for the existence of a task difficulty (or workload) homeostasis instead of or in addition to risk homeostasis? In addition to some findings against risk homeostasis theory, mainly results derived from one experimental paradigm are put forward (Fuller, 2005, 2007; Fuller et al., 2008). This paradigm involved presenting a video to participants which showed driving on a stretch of road with different speeds (changing in 5 km/h increments). The participants had to rate task difficulty, feeling of risk and subjective probability of an accident for each condition. For illustrative purposes, prototypical not real results are presented in Figure 11.

The results depicted in Figure 11 are interpreted by Fuller (Fuller, 2005, 2007; Fuller et al., 2008) as being in favour of the TCI-model: the point where estimated crash risk rises above zero is seen as equalling the point where task demand exceeds task capability (see also discussion below). In the experiments, Fuller found that 95% of the participants reported to be uncomfortable at speeds which are equal to or higher than this point. Furthermore, according to Fuller, the steady increase of feeling of risk and subjective task difficulty with increasing speed indicates that speed is indeed a strong determinant of task difficulty. This implies that vice versa it is used as a regulating mechanism by the driver (though not necessarily the only regulation mechanism, as Fuller points out).

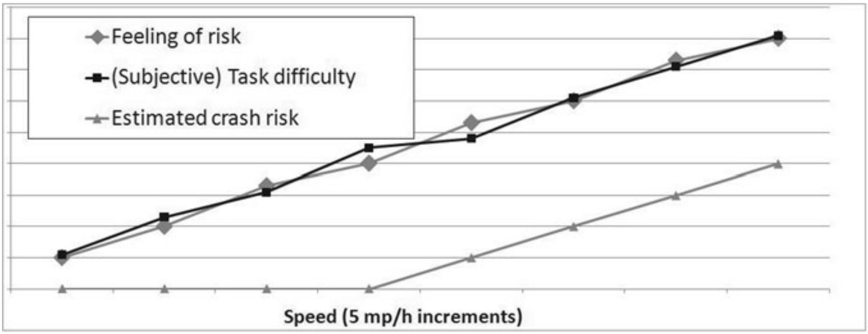


Figure 11. Prototypical results found by Fuller (2007) in support of the TCI-model.

The dissociation between feeling of risk and estimated crash risk together with the fact that participants indicated that their preferred speed is below the rise of estimated crash risk above zero (the latter is not shown in Figure 11) are both interpreted as speaking against risk homeostasis theory, at least in everyday driving. This is because according to Wilde's RHT (see chapter 2.3.4), drivers seek a certain (estimated) statistic risk as target risk and not a subjective feeling of risk. Nor do drivers seek zero statistical risk as proposed by Näätänen & Summala (1976).

However, Summala (2007) challenged these interpretations. The criticism can be summarised as follows:

- The fact that estimated task difficulty correlated almost perfectly with feeling of risk (*r* = .97 in Fuller 2005) is interpreted as being an effect of the assessment in the laboratory instead of on the road. Such conscious, rational risk ratings 'from arm chair' (Summala, 2007) do not grasp real feelings of risk and possibly task difficulty as well.
- The finding that participants reported feeling uncomfortable at the very latest at speeds beyond the point where estimated crash risk is rated as being above zero is interpreted by Summala (2007) as strong indication in favour of the zero risk model developed by Näätänen & Summala (1976) (see above) and not in favour of the TCI-model. In fact, due to the steady and linear increase of both subjective demand and feeling of risk, an exact prediction of preferred speed is impossible without the function of estimated crash risk because only the latter provides a cut-off value. Of course, as already indicated, Fuller sees the cut-off value at the point where demand exceeds capability.

The close association between feeling of risk and subjective task difficulty is interpreted by Fuller (2007) as being a result of the mechanisms involved in inferring task difficulty. Fuller (2007) speculates that drivers use ‘somatic markers’ (Damasio, 1994) of which feeling of risk might be one used to infer task difficulty. Critically, it should be mentioned that no conclusive reason is given why task difficulty should be inferred from feeling of risk and not vice versa, and why task homeostasis and not feeling of risk homeostasis should be the important mechanism in driving.

Nevertheless, the TCI-model is seen as a valuable contribution towards rural road safety. Firstly, it associates the well proven relationship between task demand, workload and performance (de Waard, 1996) with safety. Secondly, it discusses the dissociation between feeling of risk and estimated crash risk which in my opinion is indeed a weak point in Wilde’s RHT. Thirdly, it stresses the importance of task demand regulation via speed regulation. This aspect cannot be accentuated enough with speed being the single most important contributing factor on behalf of the driver to accidents with personal injuries on rural roads (Statistisches Bundesamt, 2007).

2.3.6 A Brief Overview of the Measurement of Workload

Workload can be assessed by a number of different measures and methods of which de Waard (1996) distinguishes three broad categories:

- self-report measures;
- performance measures; and
- physiological measures.

By further dividing performance measures into primary and secondary task measures and by adding the visual occlusion technique (Gelau & Krems, 2004; Krems, Keinath, Baumann, & Jahn, 2004) as an additional category, these three categories are extended to five by Johansson et al. (2004).

A decision for a specific measure or measurement technique should be based upon the criteria which were summarised by O`Donnell and Eggemeier (1986, cited in de Waard, 1996; Wickens, 1992; Zeitlin, 1995):

- Sensitivity: ability to distinguish between different levels of workload.
- Diagnosticity: ability to distinguish between different types of workload (see Wickens).
- Selectivity: ability to distinguish the result from other psychological constructs.
- Obtrusiveness: the process of measurement should not interfere with primary task performance.
- Reliability: a prerequisite of all behavioural and psychological measures.

Each criterion has to be considered for every task and every environment or experimental setting in which it is used. Deviations in the results between measures from the different measurement categories named above or between different measures within one category are likely to be attributable to differences in one or many of the aforementioned criteria and interactions between the methods themselves.

In line with de Waard (1996) a combination of several techniques was used to assess workload for the thesis at hand. Which ones were used specifically depended on the study and is discussed in the methodological chapter for each experiment. However, subjective ratings were usually collected together with primary task measures. The latter are recommended for inclusion because of their crucial importance for safety (Wierwille & Eggemeier, 1993). In addition to these measures, a secondary task was used for the driving experiments in the field (see chapter 4.4.3.8). A secondary task is a task which is not required for safe driving. The interpretation of secondary task measures is based upon the dual task paradigm (Brown, 1978; Hicks & Wierwille, 1979; Pew, 1979; Tsang & Vidulich, 2006; Verwey, 2000).

Psycho-physiological measures were not applied in the studies conducted for this thesis. This decision is amongst others based upon the discussion of electrodermal activity in chapter 2.3.4. There, electrodermal activity served as an indication of subjective risk which already reveals a general weakness of psycho-physiological variables: they are not very selective (see criteria above). Details concerning psycho-physiological measures, partly with relation to driving, are discussed in Piechulla (2006), Wagner (2000) and Manzey (1998). Visual demand was not assessed with the visual occlusion technique (see above) but with different gaze parameters where applicable (see chapter 4.4.3.9).

The measures which were used in this thesis are explained and discussed in detail in the respective chapters. An additional discussion and introduction of different measures cannot be done as part of this thesis due to the multitude of measures. Readers interested in further reading are referred to de Waard (1996), Johansson et al. (2004), and Weller et al. (2006). Additionally, an overview of several measures for self-reports and performance measures is given in Gawron (2008).

2.3.7 Motivational Target Variables: Can They be Distinguished?

So far, subjective risk and workload have been introduced as variables which could serve as motivational target variables for a homeostatic regulation of driving. However, in principle, any psychological state or feeling could serve as such a target variable. For example, Rothengatter (1988) amongst others identified ‘pleasure in driving’ as a relevant variable. Other authors propose ‘comfort’ (Summala, 2005, 2007), several motivational factors said to be either cost or benefit (Taubman-Ben-Ari, 2008) or a more general 'target feeling' (Vaa, 2007). As entire theoretical frameworks are often built around each of these different potential target variables, it seems wise to ask how these variables differ from each other.

For example, although subjective risk and workload can be clearly differentiated from a theoretical point of view (see the definitions in the respective chapters), practical application in studies tells another story. The discussion on the interpretation of electrodermal activity (see chapter 2.3.4) might serve as an example here as well as the close relationship of subjective demand and subjective feeling of risk depicted in Figure 11. When this close relationship was discussed in chapter 2.3.5, it was stated that no conclusive reason can be found for it unless one assumes that drivers do not indeed distinguish between both constructs. This would have highly relevant consequences for practical tests as it would not only diminish the effort to collect the data, but could also unify the different existing theories of driving behaviour.

However, not all authors would assent to this step (other than the ones who proposed theories built around such single constructs). For example, Groeger & Chapman (1996) and Groeger (2000) state that (ratings of) difficulty and danger can indeed be distinguished. Taking a closer look at their results, difficulty and danger were the names of factors which were extracted after principal component analysis (see also chapter 4.2.4.6) of several items. These items were used to rate filmed driving behaviour at junctions. The notion that danger and difficulty can be distinguished was based on the finding that older drivers had higher factor scores in the 'danger' factor than younger drivers, while on the other hand, neither group exhibited such differences in a factor which was termed 'demand' (Groeger, 2000; Groeger & Chapman, 1996).

However, in the article (Groeger & Chapman, 1996) it was also reported that the factor-scores of 'danger' and 'demand' for the 64 participants were significantly correlated. As the factor solution was Varimax rotated before principal component analysis, this finding is quite astonishing. In my opinion, this refutes the idea that rated danger and demand can easily be distinguished. This is also supported when taking a closer look at the items constituting the respective factors. Whereas the item 'How much risk would you have felt in that situation?' clearly supports the name 'danger', two out of the other four items constituting the factor 'danger' are far less easy to interpret. These other two items were 'How hard would you need to concentrate to drive safely in this situation?' and 'How stressful would it be to drive in this situation?'. In my view both constitute aspects of demand. Furthermore, the item 'How many accidents do you think occur at this junction?' (rating from 'none' to 'many' on a 7-point Likert scale) is part of the 'difficulty' factor but again in my view should be part of the 'danger' factor.

Thus, it must be concluded that further research is needed to establish if and how drivers distinguish between the different constructs named so far. It could well be that different constructs are active at different times or in different situations depending on the current motivation, differential variables and the current purpose of the trip. In this case, behaviour would mathematically be described as a vector in a multidimensional space of which the dimensions are represented by the strength, that is, the weight of the different target variables. Unidimensional models would then just constitute the special case of all other constructs having a weight of zero. In the case of a close relationship between two or more constructs, all would have the same weight.

2.3.8 How is the Target Defined?

In the preceding chapter it was stated that it is difficult to distinguish between the different potential target variables. However, regardless of the kind of target variable, the unifying element between all theories introduced so far is that they assume that drivers strive to maintain or reach a certain target level of the target variable. Where this target level differs from zero, reducing the target state is seen as the only way to increase safety (Wilde, 1994, 2001). This target state could be reduced by applying one or all of the following possibilities: increase the costs of risky behaviour and decrease its benefits and increase the benefits of safe behaviour and decrease its costs (Schlag, 2004).

However, a more precise definition of where exactly the target state is situated would possibly allow more differentiated countermeasures. As has been pointed out, workload is lowest and performance is best at medium levels of demand (de Waard, 1996, see Figure 8), arousal (Hebb, 1955) or subjective risk (Zuckerman, 1976, 2007), while at the same time pleasure or hedonic tone are highest at medium stimulus intensity (Berlyne, 1960, 1970; Wundt, 1874). Therefore, it might be justified to assume that drivers strive for medium levels of demand or risk and adjust speed accordingly. Unfortunately, such a medium level can only be defined in relation to the individual resources at a given time (see above).

Apter (1984) and Franken (2007) similarly point out that the optimal level of arousal depends on the respective strength of two opposing states or motives: an achievement state which is governed by avoiding anxiety and a pleasure-seeking state which is governed by seeking excitement. When in a state of excitement seeking, the individual prefers much higher levels of arousal than when in a state of anxiety avoidance. Only if averaged across several observations does a medium level of arousal result in the highest hedonic value, which – in this case – is also only medium.

Zuckerman (1976; 2007) proposes a two-motive model which is applicable to driving (see also chapter 2.2). In contrast to Apter, who sees either one or the other state as being active, Zuckerman regards both motives as being active at the same time. The two drives or motivational forces used by Zuckerman are sensation seeking and anxiety, where both are interpreted here as 'affective states' and not traits (as was the case in chapter 2.2). Positive arousal which results in an approach motivation is associated with sensation seeking. Negative arousal which results in withdrawal is associated with anxiety. Both sensation seeking and anxiety increase with subjective risk. For sensation seeking this increase follows the well-known inverse U-shaped function, whereas it follows a linear function for anxiety.

According to the graphics used by Zuckerman, the apex of sensation seeking – and thus the optimal level of arousal – is reached with maximum novelty of a stimulus or situation. This idea is influenced by the work of Berlyne (1960; 1970) who found that rated 'pleasantness' and 'interestingness' increased with the novelty of artificial graphic symbols in relation to other graphic symbols.

However, Zuckerman (1976) also notes that: 'Novelty per se, increases fear only in animals and young children. The adult human has so many ways to interpret unusual situations that no situation can be completely novel' (pp. 164, 165). Therefore, again, the optimum level of arousal is solely determined by a single parameter (subjective risk in this case) which does not allow determining its exact location or value.

Nevertheless, the relationship proposed by Zuckerman at least permits determining the end of an approach action. This point is situated slightly left of the intersection point of sensation seeking and anxiety. If approach is equated to acceleration and withdrawal is equated to deceleration, this means that speed is not increased further beyond the point where anxiety exceeds sensation seeking. The intersection point itself represents a classic approach-avoidance conflict as described by Lewin (1982).

Summing up this chapter, it can be concluded that the target value constitutes an internal value which depends on the current motivation and state of the driver in relation to objective demand and which, therefore, cannot be inferred from an objective characterization of a (driving) situation alone. It can only be inferred indirectly by observing behaviour. At this point, it must be mentioned that the presence of a comparator in which the target state is assumed to be compared to the actual perceived state of the target variable is in general criticised by some authors as 'homunculus' regardless of whether subjective risk or workload are used as target variables (Michon, 1989, cited in Rothengatter, 2002).

2.3.9 Homeostatic Regulation and Failure Thereof, Discussed in Two Examples

In the preceding chapters different theories on how drivers regulate their behaviour were introduced and discussed. Some of their assumptions are contradictory such as a zero risk versus a target risk strategy and cannot be valid at the same time for the same situation. A reconciliation of positions is only possible by assuming that each of the theories is valid only for specific situations or under specific circumstances (Fuller, 2005). Either way, the empirical evidence delivered by the authors of the theories is often sparse and the practical applicability of the theories to driving on rural roads is thus unclear.

This chapter firstly summarises the theories with respect to testable implications and then discusses the findings of selected studies with relation to the theories. This is not done so as to discuss the theories in general, which has been done at length by other authors (see references in the preceding chapters), but in order to test the relevance and applicability of the theories with respect to rural road safety. The variable of road width and curve radius were selected to summarise the current understanding and its applicability in the assessment of the different theories.

2.3.9.1 Preliminary Summary of Existing Theories With Respect to Practical Application to Rural Road Safety

As driving is a self-paced task, the drivers have a powerful means to adapt their behaviour to a changing environment. The different (motivational) theories assume that the driver follows a target state in workload (Fuller, 2005), or subjective risk, whereas in the latter case the target value is either nil (Näätänen & Summala, 1976) or somewhere above zero (Wilde, 2001). In fact, risk and arousal are hard to distinguish as was already discussed through electrodermal activity. EDA-activity seems to be related to emotional arousal (anxiety and fear) and also to (physical) workload or task difficulty. While the first interpretation is used by Wilde, the second is favoured by Fuller.

Regardless of the variable used, these theories agree that the regulation of behaviour around the target value is achieved by homeostatic processes. This implies that the parameter under consideration varies within a small range around the target value, where the range can be defined as the target value plus/minus the 'just noticeable difference' (Wilde, 2001). Unfortunately, research in traffic is not often carried out to explicitly test theoretical assumptions. More often research is carried out to compare different designs by comparing performance indicators such as speed or accidents without explicitly testing the underlying psychological mechanisms. Nevertheless, in order to incorporate these studies into this thesis, the performance indicators used must fulfil certain criteria.

'Performance' can be defined in terms of safety and in terms of mobility, although ideally both should be achieved at the same time. However, mobility is still associated today with a certain amount of risk and thus increased mobility might increase the total number of accidents. Therefore, optimal performance is situated at the point where mobility is highest and risk is still 'acceptable', both on the individual and the aggregated societal level. If risk was referenced to unit time, this amount of risk could be called 'target risk' in the sense of Wilde. A parameter which takes into account mobility is the accident-rate or the accident cost rate (FGSV 2003). As a change in road width and curve radius might also change the amount of cars on these roads (and vice versa if surveyed by the road authorities), it is vital to use these accident parameters instead of merely the number of accidents.

The importance of exposure for the development of accidents also becomes evident when considering that average annual daily traffic (AADT) is the most important – or even the only parameter – in a number of accident prediction models (Reurings et al., 2005). Furthermore, accidents are rare events from a statistical perspective and are therefore not suited to evaluate driving strategies at the individual level (discussion of accident properties in Weller et al., 2006). In addition, no accident data is available when assessing the safety effects of new road or vehicle designs and in this case performance parameters at the individual level have to be used as proxy variables for future accident occurrence.

This implies that the relationship between driving behaviour and accident occurrence is known. Unfortunately, such information is rare. Ongoing research projects (e.g. GIDAS) try to increase understanding of both the effects and causes of accidents by in-depth on-the-spot accident analysis. Reichart (2001) used another approach and successfully transferred knowledge on human reliability and error likelihood to car driving. The connecting link between driver and driving behaviour on the one hand and accident occurrence on the other were driver and driving errors which were also defined by Reichart (2001). A method partly based on this approach was successfully implemented in an assessment procedure for advanced driver assistance systems (ADAS) (Glaser, Waschulewski, & Schmid, 2005; Nirschl, Böttcher, Schlag, & Weller, 2004; Weller, Schlag, & Nirschl, 2006).

A discussion of the theories named so far requires that data concerning workload and subjective risk are reported in addition to the appropriate performance parameters. Preferably, these performance parameters should also be linked to accident data at the experimental road section. It is the combination of all variables which allows a meaningful validation of the theories which claim to be able to explain accident occurrence based on psychological variables.

To summarise the preceding paragraphs, the following measures must be available to assess the validity of motivational theories with respect to selected road features:

- objective characteristics of the situation as indicator of demand;
- accident (cost) rates at the aggregated level;
- driving behaviour at the individual level, where speed is the most important variable as it is a direct indicator of the self-paced nature of the driving task; and
- subjective risk and workload at the individual level.

The next two chapters assess the quality of available data for the two examples of road width and curve radius.

2.3.9.2 Existing Findings Concerning the Effect of Road Width Discussed With Relation to the Existing Theories

Road or lane width[5] was selected because of several reasons. Firstly, it can easily be measured and thus allows assessment of objective demand. Secondly, road width not only affects the costs of a road but is also a central variable in self-explaining road categorisation (Matena et al., 2007). Therefore, it was assumed that not only its effects on accidents but also on behaviour, workload, and/or subjective risk should be well documented.

As the central independent variables in the theories are objective demand or risk, how does road width affect these parameters? In short, increasing road width should result in a decrease of objective demand and risk provided that all other factors remain unchanged. This is because the time-to-line-crossing (TLC, see chapter 2.4.4) increases and leaves higher time margins for corrective action (see also Summala, 1996). On the other hand, increasing speed on a given cross-section will increase demand. This was shown by Godthelp, Milgram & Blaauw (1984) who found an increasing need for information (operationalised by occlusion time) with increasing speed on straight roads in real road driving. In the preceding chapters it was stated that drivers are supposed to adapt their behaviour to changes in the environment. A

---

[5] Road width includes hard shoulder width and lane width in contrast to carriage width which does not include hard shoulder width. Lane width in turn includes all adjacent lines.

likely change in behaviour as a consequence of decreasing demand would be an increase in speed. According to homeostatic theories this increase in speed because of decreasing demand would result in essentially the same value of the target variable (see above). Furthermore, if this strategy was successful, the outcome of performance should also be more or less constant. This means that accident (cost) rates should be comparable between roads of different widths.

Despite the central role of road width for road design, studies dealing with the effect of road width on accidents are astonishingly scarce. The reason is most likely that it is difficult to isolate the influence of road width because changes in road or lane width coincide with other changes of the road layout (Lamm, Psarianos, & Mailaender, 1999). This paragraph summarises evidence on the relationship from different sources.

Cohen (1997) refers to a Swiss study (Dietrich et al., 1983, cited in Cohen, 1997) showing a U-shaped function of accident-rate to road width. However, this Swiss study comprised road widths ranging from eight to 14 metres and the resulting accident-rate might therefore have also been influenced by the number of lanes. The number of lanes is not reported by Cohen (1997) but is known to have an influence on the number of accidents itself (Elvik & Vaa, 2004).

Becher et al (2006) cite a report by Brannolte et al. (1993) which shows a higher accident-rate for narrow two-lane roads in comparison to wider two-lane roads (2 × 3.75 m and 2 × 3.50 m versus 2 × 3.25 m and 2 × 3.00 m) but also a U-shaped function when accident cost rates are considered (same cross-sections). Such a pattern is usually caused by higher speeds which cause higher accident costs due to the higher kinetic energy involved (Aarts & Schagen, 2006).

Lamm, Psarianos & Mailaender (1999) summarise numerous empirical studies regarding the relationship between roadway width and accidents, accident-rate or accident cost rate. All studies cited by Lamm et al. (1999) showed that either the number of accidents or the accident-rate decreases with increasing lane width. With respect to the accident cost rate, Lamm et al. (1999) report some controversial findings and conclude that additional research is needed in order to arrive at reliable conclusions.

Elvik & Vaa (2004) also summarise several studies, some of which are included in Lamm et al. (1999). Elvik & Vaa (2004) made a distinction between increasing road width (which includes shoulder width) and increasing lane width. The results for increasing road width show a decline in the number of accidents for rural roads. The results concerning lane width are inconsistent and range from a decrease to an increase in the number of accidents.

Vogt and Bared (1998) developed an accident prediction algorithm for segments of two-lane rural roads which is used as the base model in the Interactive Highway Safety Design Model (IHSDM). The application of this algorithm results in a decrease of 8.1% in the predicted number of accidents with every foot (= 0.305 m) increase in lane width. The base model of the IHSDM is calculated with fixed values for geometry. For example, the value for lane width is set to 3.6 m. Only then it is adjusted with accident modification factors (AMFs) which take into account the actual values of the geometric elements under consideration (Harwood, Council, Hauer, Hughes, & Vogt, 2000). These AMFs are based on the judgement of an expert panel which in turn took into account various research results. Similar to the work of Vogt and Bared (1998), the AMFs mirror an increase in the number of accidents with decreasing lane width. An increasing lane width beyond the base value of 3.6 m presumably does not have an influence on the accident number as these widths do not result in an adjustment of the AMF; neither do values below 2.7 m (Harwood et al., 2000).

It is important, that Harwood et al. (2000) point out that the effect of lane width on accident occurrence depends on the AADT: there is hardly any effect for AADT < 2000 vehicles/day. This might also explain some of the inconsistencies in the findings between the studies which were summarised by the authors named in the preceding paragraph. Despite these inconsistencies, it can be assumed that the accident-rate increases with decreasing road width. This indicates that drivers were not capable of adapting their behaviour successfully to the higher demand associated with decreasing road width. This should be mirrored by a zero or insufficient reduction in speed on narrow roads or lanes in comparison to wider roads or lanes and/or an increase in workload.

In contrast to the effects of road width on accidents, more evidence can be found concerning the influence of road width on speed. Several reviews on this topic, which all summarise several other studies, agree that in general speed increases with increasing lane width (Martens, Comte, & Kaptein, 1997; Matena et al., 2006; OECD, 1990). However, the influence of road width on speed is quite low compared to the influence of the radius in curves for speed (Lippold, 1997), compared to the influence of sight distance on local, urban streets (York, Bradbury, Reid, Ewings, & Paradise, 2007) or compared to the posted speed limit (Fitzpatrick, Miaou, Brewer, Carlson, & Wooldridge, 2005). Further, Lippold (1997) points out that differences in speed might only be found when extreme values between two classes of road width are compared. The limit between these two classes might be situated at a value of 6 to 6.5 metres (Lippold, 1997).

So far, the increased accident-rate on narrow roads despite slower speeds indicates that the adaptation processes by the drivers were not enough to compensate for the increasing demand. Of course, this relationship could be formulated more positively as is done in the OECD report on behavioural adaptation (OECD, 1990) which states that despite an increase in speed with increasing road width '… the evidence suggests that there is a substantial net benefit associated with wide pavement lanes and shoulders' (p. 41).

As the focus of this chapter is on the underlying mechanisms behind driving behaviour and resulting accident occurrence then the question is what is the effect of road width on psychological parameters? As would be thought, there are far fewer studies assessing workload or subjective risk compared to studies solely measuring speed. The following studies were selected to report these findings: Lewis-Evans & Charlton (2006), de Waard et al. (1995), Cohen (1997), Godley, Triggs & Fildes (2004), and Steyvers & de Waard (2000).

Lewis-Evans & Charlton (2006) used a desktop driving simulator to assess speed on roads of different width and found that speeds were lower for a narrow road (road width 8.6 m, lane width 3.0 m) in comparison to two wider roads (road width 10.6 m and 13.6 m and lane widths 3.6 m and 4.6 m). Although speed was higher for the widest road in comparison to the medium width road, this difference was not significant. Ratings for risk, difficulty and accident risk were all higher for the narrow road in comparison to the medium and wide road. Interestingly, the participants did not show any open awareness of the differences in road width but named several other aspects as determining factors. Although the results are interpreted in favour of the zero risk theory by Summala (1988), the authors also state that the non-significant differences in speed between the widest and medium road could be the result of a ceiling effect.

De Waard et al. (1995) examined the effect of 'intermittent chipped road markings' (presumably rumble strips) on two-lane rural roads on driving behaviour and on different psycho-physiological parameters. The rumble strips made driving uncomfortable if crossed and thus reduced the 'comfort zone' and implicitly lane width from 2.70 to 2.25 metres. The experiments were conducted on real roads similar to the roads used in the studies conducted by Steyvers (see chapter 4.2.2). The different road layouts resulted in significant lower speeds for the experimental condition, whereas an additional environmental effect of curved forest versus straight moor section was also significant. The differences in mean speed between control and experimental section were 3 km/h (forest) and 1.5 km/h (moor). Given that the dimension of a design vehicle used in road planning in Germany is 2.5 m (Lamm et al., 1999) and that the lane width of the experimental condition is even below this value, these differences in speed are quite small. Different cardiovascular parameters which were used as indicators of workload showed higher workload for the experimental road sections. In addition to psycho-physiological data, the road environment construct list (RECL, see chapter 4.2.2) was applied. Hedonic value was lower for the experimental condition while the other two factors did not show experimental effects.

Cohen (1997) investigated the effect of road width on spare capacity (as the inverse of workload). Spare capacity was measured as latency time between auditory presentation of a stimulus and the reaction to this stimulus. The experiments were carried out on two-lane rural roads of different carriageway widths (6-9 metres).

Workload was lowest on medium width carriageways (8 metres). This U-shaped function of workload with demand shows that drivers were not able to keep workload at a medium level independent of demand as is proposed by workload homeostasis. Astonishingly, the speeds driven were also highest for the medium width carriageways. Combining both findings it might be that demand does not rise linearly with road width but that there is indeed an optimal road width for human drivers which is at medium levels and not at the widest width. An alternative explanation might be that drivers paid more attention to driving when increasing speed. This could have increased capacity, which is the upper level of resources (Young & Stanton, 2002), and thus led to lower reaction times. To investigate such effects, the drivers would have to drive for a longer time.

Godley, Triggs & Fildes (2004) investigated the effect of perceptual road width on speed and workload in a simulator. Although the finding was not consistent throughout all experimental conditions, in general, mean speed was lower on perceptual narrow lanes. Workload, assessed with the NASA TLX, subjective risk, and steering effort all increased with decreased lane width, although again the changes were not significant.

However, perceptually reducing the road width might also lead to opposite effects on speed as was shown in a study by Steyvers & de Waard (2000) on Dutch rural roads with few ADT. In two experiments, one with stationary video recordings and the second with an equipped vehicle, the authors compared lined and unlined roads. In the lined situation the position of the cars shifted to the middle without moving into the adjacent lane. It was thus seen as a successful strategy to counteract the high amount of run-off-the-road accidents on rural roads. However, in this study (Steyvers & de Waard, 2000) the lined situation resulted in lower workload ratings, lower steering effort, lower subjective risk ratings and higher speeds than the unlined situation. In this case, the lines were obviously perceived as guiding elements rather than as narrowing the road. Therefore, the specific conditions of the whole environmental situation have to be taken into account when applying perceptual countermeasures. The same conclusion can be drawn from inconsistencies concerning the effects of markings on accident figures reported by the OECD (1990). In addition, while effects (positive or negative) are to be expected for drivers unfamiliar with the road, the effect of road-markings wears off quite rapidly for drivers familiar with the road (e.g., Cavallo & Cohen, 2001; Elvik & Vaa, 2004; Smiley, 1999).

Although none of these studies refer to accident data, the findings are seemingly in line with the conclusions drawn above concerning the relationship between road width, speed and accidents. This means that despite an adaptation of speed, subjective demand and subjective risk were higher on the narrow roads, indicating insufficient speed adaptation. The higher accident-rates on narrow roads could be explained by insufficient or unsuccessful speed adaptation which results in higher workload or risk.

However, the studies which assessed workload and risk with subjective ratings have a severe shortcoming: this assessment method might not have taken into account the influence of speed on the ratings. That is to say that participants rate the situation as they see it without the influence of speed instead of the situation as it is with the given speed. Therefore, studies which have assessed workload and risk with psycho-physiological data (de Waard et al., 1995) or with a secondary task (Cohen, 1997) might be more valuable. However, in these studies it is difficult to distinguish between cause and effect, although an increase in workload with demand despite reduced speed would be interpreted as being safety critical by the theories.

2.3.9.3 Existing Findings Concerning the Effect of Curve Radius Discussed With Relation to the Existing Theories

Similar to the influence of road width, the influence of curve radius on accident parameters, performance indicators, and workload or risk will be summarised. The parameter curve radius differs from road width because curve radii show a much higher variation due to the topography and because they require an anticipated adjustment by the driver. Over 30% of accidents with personal injury occur in curves and thus curves are singularly the most notorious design element when it comes to severe accident occurrence (Ellinghaus & Steinbrecher, 2003; Statistisches Bundesamt, 2007). Therefore, it can be assumed that at the very least the relationships between radius and accidents and between radius and speed should be well established.

With respect to the relationship between radius and accident-rate, Elvik & Vaa (2004) analysed several studies conducted worldwide and found a reduction in accident-rate with an increase in radius. This relationship is valid up to a radius of approximately 1000 to 2000 metres from which point a further increase does not lead to a further decrease in the accident-rate. Similar relationships are used in the IHSDM (Harwood et al., 2000) which is also summarised in the report by Mallschützke et al. (2006). Further studies which were summarised in the report by Dietze et al. (2005) found similar relationships: both an increase of the accident-rate with decreasing radius and a less pronounced relationship for radii above 400 metres.

In addition to the ‘simple’ radius, the curvature change rate (CCR) can be used as a characterising parameter of the curve. Using CCR has the advantage that the length of the curve is also included (see chapter 4.3.3.12 and Appendix A 1.4). Lamm et al. (1999) found an increase in the accident-rate with increasing CCR. For practical applicability to road design, not only should the characteristics of the single curve be taken into account when estimating an effect on accident occurrence but also the characteristics of the sequence of curves and the intermittent tangents must be considered (Dilling, 1973; FGSV, 1995; Fitzpatrick, Wooldridge et al., 2000; Lamm et al., 2007; Lippold, 1997).

The relationship between radius and speed is well established, although different authors have found different mathematical functions. In general, the 85th percentile of speed (V85) increases with increasing radius or curvature (FGSV, 1995; Fitzpatrick, Elefteriadou et al., 2000; Lamm et al., 2007; Lippold, 1997). In the review conducted by Dietze et al. (2005), the authors further show that the influence of radii beyond 350 m declines and, by referring to RAS-L (FGSV, 1995) that road width in curves also must be considered.

The effect of radius on workload or risk is less documented than the effect of radius on accident occurrence or speed. Therefore, similar to road width, several publications which deal with this subject must be introduced in more detail to infer appropriate conclusions for the validity of the assumptions made by the theories. The studies used are: Backs, Lenneman, Wetzel, & Green (2003), Fitzpatrick et al. (2000), Tsimhini & Green (1999), Richter, Wagner, Heger & Weise (1998), Wagner (2000), Van Winsum & Godthelp (1996), and Messer (1980). Additionally, the works of Hoyos & Kastner (1987) and Hoyos (1988) were considered. Despite the fact that these works do not deal with curve radius, they did test workload homeostasis and are thus regarded as important in the present context.

As has been pointed out, an important aspect in motivational theories is the self-paced nature of the driving task which allows an adjustment of speed to differing demand. This adjustment in turn might lead to a more or less constant level of workload or risk. The following three studies (Backs et al., 2003; Fitzpatrick, Wooldridge et al., 2000; Tsimhoni & Green, 1999) did not allow drivers to adjust their speed and thus strictly speaking do not allow a statement concerning the validity of the theories. However, they allow the determination of actual workload as a result of radius without speed adjustment. This is an important aspect because the findings can subsequently be used to interpret results found with speed adaptation. The three studies mainly used visual occlusion, electrocardiac measures and subjective ratings in simulated and test track environments to assess curve demand. A detailed summary of the results of Fitzpatrick et al. (2000) can be found in Weller et al. (2006). While each of the studies differed with respect to the investigated conditions and parameters, speed was usually set at around 70 km/h and the investigated curves had a radius of 582, 291, 194 and 146 metres each with deflection angles of 20, 45 and 90 degrees[6].

Performance, measured by different steering and lane keeping parameters worsened with decreasing radius (and with visual occlusion), whereby the decline

---

[6] The deflection angle is the angle between approach and departure tangent of a curve. For a given radius the deflection angle is thus also a measure of curve length (see Appendix A 1.4). A deflection angle of 90 degrees allows the driver to make a full right or left turn.

followed a near linear function. The influence of radius was much stronger than the influence of deflection angle. The results also showed a marked increase of workload measured by occlusion parameters and by subjective ratings with decreasing radius. Tsimhoni & Green (1999) further showed that the workload profile increased before the curve and reached its peak immediately before entering the curve at the end of the approach tangent, independent of curve radius. The results for the psycho-physiological measures as applied by Backs et al. (2003) were less conclusive with respect to radius. Nevertheless, the studies indicate that demand and workload in fact increase with decreasing radius. Thus, the regulation of speed such as a decrease in speed with decreasing radius could indeed be explained as a function of demand or workload.

The study by Shinar, Rockwell & Malecki (1980) showed that perceptual countermeasures (achieved through different kinds of markings and a sign labelled 'deceptive curve') could significantly reduce speed in rural road curves. Together with the findings of Godley, Triggs & Fildes (2004) for road width (see above), this again shows the importance of perceived task characteristics compared to actual, objective task characteristics.

Richter, Wagner, Heger & Weise (1998) and Wagner (2000) investigated the influence of the curvature change rate of entire road sections on mental workload. The length of the investigated sections was between three and five kilometres. Mental workload was assessed by several psycho-physiological indicators and with subjective ratings. The study was conducted on real rural roads with 31 participants and thus involved considerable effort in data collection. Speed, which was recorded during driving, could freely be chosen by the participants. The analysis of the data showed that average speed was highest for the lowest CCR road and lowest for the highest CCR road, but showed no systematic variation between roads of different medium CCR levels. The results for the subjective ratings and several psycho-physiological parameters exhibited similar effects. In contrast to these results, blink rate increased with decreasing CCR across all CCR values. This indicates a decreasing demand with decreasing CCR.

As all parameters were averaged across each experimental road section, it is not possible to reference the parameters to characteristics of single elements (Wagner, 2000). The fact that systematic differences were found nevertheless, at least between the highest and lowest CCR road section, indicates that if homeostasis existed, it was not successfully achieved by drivers. As all psycho-physiological values are referenced to time (e.g. beats per minute), this also applies to the claims made by Wilde (1994; 2001).

Van Winsum & Godthelp (1996) examined the relationship between curve radius, speed and steering errors in simulated curve driving. Steering error is defined as deviant steering wheel angle from required steering wheel angle and is calculated as integral over time. The required steering wheel angle in turn is defined by the curve radius and the speed driven. To compensate for steering errors due to smaller radii or less experience, drivers can reduce speed. Summing up these characteristics, it might thus be acceptable to treat steering error as a performance measure of workload.

Van Winsum & Godthelp (1996) found that steering errors increased with smaller radii despite the drivers decreasing speed at the same time. At first sight, this indicates that the drivers did not sufficiently reduce speed in curves with smaller radii. However, the steering error ratio (calculated as steering error divided by required steering wheel angle) and minimum TLC remained constant, independent of radius. This finding is interpreted by the authors as supporting evidence in favour of driving being a task which is controlled by time margins as is assumed by Summala (1997). Whether time margins have to be interpreted as indicative of subjective risk or workload as in the case of visual occlusion must remain open at this stage.

The work of Messer (1980) initially offers a clear relationship between workload and geometric curve characteristics with an increase of workload with both the degree of curvature and the deflection angle. However, Messer did not use measured workload but rather the expert ratings of 21 highway design engineers as the basis with a subsequent extrapolation based on the literature to derive the so-called workload potential ratings.

An evaluation of workload homeostasis can also be seen in the driving studies conducted by Hoyos & Kastner (1987) and Hoyos (1988). By taking into account the transactional stress model of Lazarus (e.g. 1991) and Wilde's risk homeostasis theory, the authors assumed that drivers regulate speed in order to keep a 'balance between stress and strain, a 'dynamic' sense of equilibrium' (Hoyos, 1988, p. 574) and 'to keep workload constant on an individually acceptable level' (Hoyos & Kastner, 1987, p. 50 own translation).

The authors not only tested variations of a single design element, but also the reaction to different real world traffic situations. The demand associated with these situations was assessed with task analysis along various dimensions. Strain was assessed with subjective ratings along the dimensions 'duration of strain', 'intensity of strain' and 'controllability of the situation', several psycho-physiological indicators and several CAN-bus related parameters, of which speed was one. Regarding the correlation to objective demand, the authors found positive correlations with subjective ratings of strain, negative correlations with average speed, and either positive or no significant correlations with psycho-physiological parameters. These results could be interpreted in favour of a workload homeostatic regulation of driving. However, the 'target' state as indicated by the results of the psycho-physiological parameters was not fully achieved despite lower speeds.

2.3.9.4 Conclusions Derived From the Two Examples With Respect to the Theories

The interpretation of the effect of radius on speed and workload, or risk, with respect to the theories is comparable to what has been found for lane width: the findings do not allow a final validation of any of the theories. Despite the fact that some indications were found of adaptive regulation in order to keep workload or risk constant, a homeostatic balance was never fully achieved. One reason must be seen in the data quality: as Hoyos & Kastner (1987) state, workload homeostasis is difficult to test in its entirety because it is regarded as a circular process which would require continuous assessment of the variables rather than intermittent assessment as was often done in the past (see above). Therefore, from this chapter it must be concluded that there is a considerable lack of data which could conclusively be used to support or refute any of the theories. A study is urgently needed which addresses this gap. Such a study should continuously monitor workload or subjective risk together with driver and driving behaviour as well as accident data of the situation under investigation. This thesis aims at bridging this gap.

2.4 Perception and Information-Processing

Depending on the validity of the theories discussed in the preceding chapter, the driver needs information either to determine the actual level of the target variable for the situation or, more generally, to adapt behaviour to constantly changing situational demand. This information is picked up visually with the visual sense being the predominant source of information in driving (Sivak, 1996). It can be conveyed to the driver via single cues (Posner, 1980) or signals ('Signale', Hacker, 2005). However, Michon (1985) pointed out that it might be difficult to determine which element in the driving situation acts as discriminative stimuli to determine the actual level of risk. By introducing the concept of affordances (Gibson, 1986) which are defined for an entire situation this problem can be circumvented.

This chapter distinguishes between cognitive models for which the model of Rumar (1985) serves as a framework and the direct approach to perception which was founded by Gibson (1986). Cognitive models comprise constructs such as mental models, attention or situation awareness (Baumann & Krems, 2007). Situation awareness (Endsley, 1995; Endsley & Garland, 2000) is not further discussed here because it is not regarded as being particularly relevant for driving on rural roads in free flowing traffic conditions. Besides the sources named above, a summary of situation awareness can also be found in Weller et al. (2006). As part of the direct approach to perception, perceptual invariants such as Tau or TTC and TLC as applied to driving are introduced together with the concept of affordances.

2.4.1 A Framework for Cognitive Models: Rumar's Model

Rumar (1985) provides a general framework of cognitive processes which can be used for driving and which is helpful in explaining the processes described further below (see Figure 12).

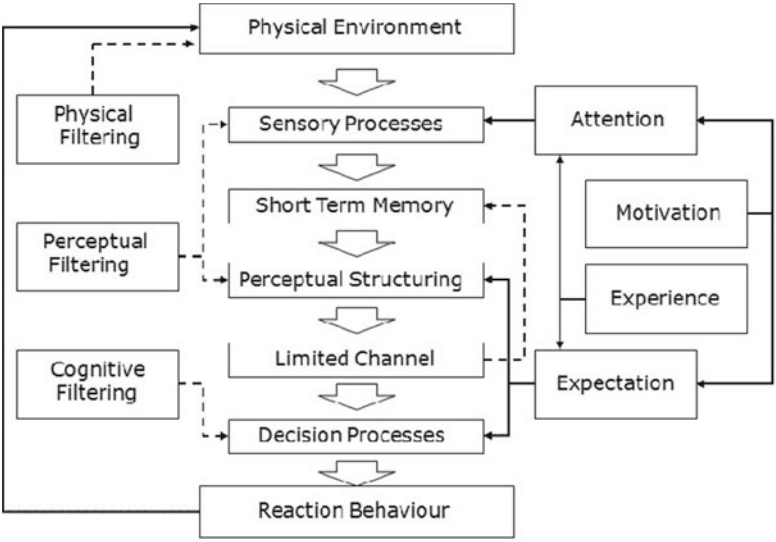


Figure 12. Information-acquisition and processing model of driver behaviour (Rumar, 1985).

Rumar's model (1985) incorporates several important aspects:

- the notion of filters, that is, that in spite of the fact that information might be objectively available it does not necessarily have to be used;
- a co-existence of environmental stimulation and cognitive processing, representing top-down and bottom-up processes in perception and information-processing (Müsseler, 2002; Zimbardo & Gerrig, 2004); and
- the interaction between different cognitive and emotional processes such as expectation, experience, motivation and attention.

The three filters named in Rumar's model (1985) are physical, perceptual and cognitive filtering. Physical filtering means that information cannot be perceived because it is concealed by another object, for example the branches of a tree. Perceptual filtering refers to the physiological limitations of the human senses. As the visual sense is the most important one in driving, perceptual filtering is determined in driving by the characteristics and limitations of the eye as the organ of vision. Some important aspects are as follows (Bruce et al., 1996; Schlag et al., 2009):

- The human eye needs time to adapt to different light conditions. The time needed for rods and cones to adapt from brightness to darkness is longer than vice versa and might be up to 30 minutes for rods (von Campenhausen, 1993). This is relevant when entering tunnels or tree-lined roads in daylight.
- The human eye needs time to accommodate from near to far and vice versa. This accommodation is relevant when drivers direct their attention from inside the car (e.g. the speedometer) to outside the car. Accommodation is faster from near to far than vice versa.
- The human eye is only sensitive for light of a very narrow bandwidth and high contrasts. This has to be taken into account when presenting information to the driver.
- Human perception depends on the context and is relative to other stimuli as shown by psychophysics (Weber, Fechner, Stevens, overview in Goldstein, 2005).
- Foveal vision is very restricted (Levi, 1999) and identification of objects often requires foveal fixation on these objects.
- The useful field of view or useful field of vision (UFOV) (Rantanen & Goldberg, 1999) changes with demand (Miura, 1990; Recarte & Nunes, 2000; Rogé, Otmani, Pébayle, & Muzet, 2008) and increasing speed (Land & Horwood, 1995). This is particularly relevant, as a decrease in the visual field and thus peripheral vision significantly reduces correct speed estimation (Cavallo & Cohen, 2001). For a discussion of the term ‘useful field of view’ and related concepts, see Weller et al. (2006), Weller & Geertsema (2008) and Crundall, Underwood & Chapman (1999).

The third filter in Rumar’s model is the cognitive filter. Cognitive filters represent the limitations of human information-processing subsumed as top-down processing. Important aspects of top-down processing are expectations which guide attention and thus facilitate or hinder the selection of relevant information. Some aspects of expectations and mental models important for driving are summarised in the next chapters. However, attention is not only directed by expectations but also in a bottom-up fashion by stimulus characteristics. An additional chapter is dedicated to attention because of its relevance to driving. The coexisting influence of both bottom-up and top-down mechanisms to direct attention explains why Rumar states that ‘sometimes the border between perceptual and cognitive filtering is unclear’ (Rumar, 1985, p. 159).

2.4.2 Expectations and Mental Models

Top-down processes basically represent 'expectations'. Expectations are derived from higher-order representations of reality in memory. The representation activated depends on the perceived similarity of the actual situation with the characteristics of the situation stored in memory. These higher-order representations were given different names during the development of psychological science. They will be introduced first and then their relevance for rural road safety will be discussed. This thesis focuses on schemata, scripts and mental models.

The construct schema was originally introduced by Bartlett in 1932 (cited in Neuschatz, Lampinen, & Preston, 2002, p. 687):

> Schemas are knowledge structures. In particular, they are organised collections of information that are stored in long-term memory, are quickly accessible, and are flexible in their use and application (Hastie, 1981). Schemas guide us as we perceive the world with which we interact. They tell us what is important and what things deserve our attention. They serve as a basis for searching memory and reconstructing it.

The role of schemata is visualised in the perceptual cycle developed by Neisser (1976) (Figure 13). Schemata help to direct our attention and exploratory actions towards the information we regard as important. They themselves are modified by the sampling of perceived object information resulting from these exploratory actions.

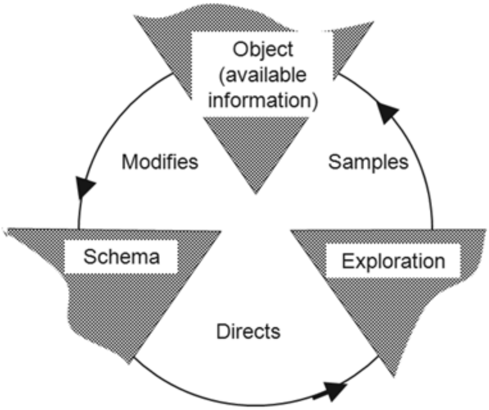


Figure 13. The role of schemata in the perceptual cycle developed by Neisser (1976).

The original schema construct was expanded to include the chain of events in time and was termed scripts by Schank & Abelson (1977). 'Event schemas, or scripts, are defined as spatially-temporally organised sequences that specify the actions, actors, and props most likely and least likely to occur during any given instantiation of an event' (Hudson & Fivush, 1992, p. 483). Missing information can be inferred with the use of scripts as long as the situation is prototypical enough for a valid script to have been developed by preceding similar situations.

Associated with scripts are the concepts of habits and routines. Aarts & Dijksterhuis define habits as 'associations between goals … and behavioural responses' and further as 'automatic behavioural responses' (Aarts & Dijksterhuis, 2000, p. 76). Whereas habits are seen to develop due to extensive iteration, the term routine is used in decision situations as the dominant solution and thus does not necessarily rely on repetition. Nevertheless, habits might be the reason why the routine is activated (Betsch, 2005).

The term mental model has become increasingly popular in human-machine interaction although it was originally used in physical science to describe the developmental process of theories (Brewer, 2002). The term is employed further in educational and developmental psychology (see Derry, 1996) and especially in the field of understanding reasoning (Johnson-Laird, 1983, reprint 1990). Mental models are essentially equal to the German concept of 'Operative Abbildsysteme' (OAS) (Hacker, 2005). Depending on their field of interest, definitions again emphasise different aspects. In summing up the commonality of these approaches, Brewer (2002) gives the following definition:

> A mental model is a form of mental representation for mechanical-causal domains that affords explanations for these domains. (…) The information in the mental model has an analogical relation with the external world: the structure of the mental representation corresponds to the structure of the world. This analogical relation allows the mental model to make successful predictions about events in the world. (pp. 5–6)

When discussing the location in memory where mental models are stored (or developed), Brewer (2002) states that the term is applied both to knowledge in long-term memory as well as to temporary specific processes in short-term memory. A definition applied to human factors research is given by Wilson & Rutherford (1989, p. 619): '… a mental model is a representation formed by a user of a system and/or task, based on previous experience as well as current observation, which provides most (if not all) of their subsequent system understanding and consequently dictates the level of task performance'.

While the distinction between schemata and scripts is self-explanatory, the distinction between both concepts and mental models is less obvious. Endsley (2000) sees schemata and scripts as being 'associated' (p. 16) with mental models. This could be interpreted as meaning that they are closely related although not necessarily a part of mental models. Other authors (Brewer, 2002; Derry, 1996) differentiate the concepts according to their stability in memory. Schemata are seen as more stable in nature whereas mental models are developed at the time of input by incorporating already existing schemata when a current situation requires them. While not mentioning schemata in the sense described above, Johnson-Laird (1983, reprint 1990) criticises the script concept by stating that it is applicable to prototypical situations only and thus is not applicable at all in the author's concept of mental models.

Although terms may differ between authors, all authors agree that internal representations (as the generic term for schemata, scripts, habits, routines and mental models) help to increase efficiency and effectiveness in human behaviour. The reasons are manifold. Firstly, they are simplified in comparison to nature. Secondly, their use is automatic rather than conscious and therefore needs fewer resources in working memory. Finally, they guide attention to relevant stimuli and therefore require less attention.

The measurement of internal representations is necessarily complex and time consuming when higher-order mental processes are involved. Often the term knowledge-eliciting is used to describe the process. Olson & Biolsi (1991) distinguish between direct and indirect methods. Direct methods mainly rely on verbal response from the participant. They include different interview techniques, 'thinking-out-loud' protocols or the observation of task performance which is interrupted at critical steps which is similar to the freezing technique used to assess situation awareness (Endsley & Garland, 2000). Indirect methods are mainly based on the result of proximity judgments between elements regarded as important. These techniques are multidimensional scaling (MDS), repertory grid or hierarchical cluster analysis. The advantages and disadvantages of the different procedures in the field of team or shared mental models are discussed by Langan-Fox et al. (2000). Finally, behaviour can be compared in different situations which are assumed to activate different mental models. Differences in behaviour can then be attributed to different mental models.

Despite their undoubted value, the effect of internal representations and the expectations derived from these representations does not necessarily have to be positive. Internal representations can be responsible for a variety of faulty actions themselves (Hacker, 2005; Norman, 1981; Reason, 1990), mainly as rule-based errors at the rule-based level (Räsänen & Summala, 2000; Reason, 1990). Events not expected dramatically increase brake reaction time (Green, 2000). Similarly, unexpected changes in the right of way increase the probability of errors especially when driving in seemingly familiar terrain (Martens & Fox, 2007). Such examples explain why expectations are a central aspect in the law of rare events proposed by Elvik (2006) as one fundamental law of accident causation.

One reason behind the negative effects is that misleading or missing information (see above: 'physical filters'), and missing or faulty use of objectively available information (see also the error-prevention oriented error classification developed by Hacker, 2005) can result in the activation of inappropriate mental models. In addition, the mental model can be faulty itself.

Because of the bottom-up activation of internal representations, misleading cues in the environment play a major role in error causation when driving. The characteristics of rural roads such as low standardization and subsequent high variation in potential behaviours make them especially conducive to errors. However, these very characteristics allow the prevention of such errors provided that the design makes use of the concepts described above.

An example of such application in road design is the self-explaining road concept (Theeuwes, 2000; Theeuwes & Godthelp, 1995) and the studies which were conducted in order to develop a self-explaining road categorisation (Riemersma, 1988; Theeuwes, 1998). Details with regard to its current state of implementation in Europe are described in Matena et al. (2007) and Hartkopf & Weber (2005).

More implicitly, expectations and mental models are successfully implemented in the engineering concept of 'consistency'. With respect to expectations, consistency means that drivers expect the following road section to be geometrically comparable to the preceding road section unless indicated by some environmental cue. Accidents often occur when the driver's expectations do not match the road situation, that is, the road is not consistent (Cafiso, La Cava, & Montella, 2007; Lamm et al., 1999). However, recently, it was found that consistency with respect to the number of road bends is not per se a protective factor for accident occurrence (Haynes et al., 2008).

Besides being used in design guidelines for rural roads (e.g. RAS-L: FGSV, 1995), consistency criteria can also be used to assess the safety level of rural roads (Cafiso et al., 2007; Lamm et al., 2007) and subsequently be applied to the development of driver assistance systems for rural roads (Ebersbach, 2006; Ebersbach & Mayser, 2004; Schwarz & Schlichter, 2004).

2.4.3 Attention

In chapter 2.3.5 workload-models were discussed and the term 'resources' was introduced together with several theories which are used to explain phenomena usually associated with attention. Therefore, the current chapter gives only a short overview of concepts uniquely used for attention, together with the appropriate definitions and brief implications for road safety.

As resources are limited and yet the environment contains innumerable stimuli, the resources must be allocated to the stimuli or the information which is relevant to driving. On the other hand, resources have to be shielded from those stimuli which are not relevant or have already been investigated (IOR paradigm, review in Klein, 2000). This allocation of resources is described by the construct attention. Luck & Vecera (2002) thus define attention as 'restricting cognitive processes to a subset of the available information in order to improve the speed or accuracy of the cognitive processes' (p. 238).

Whether attention in this sense can be viewed as a 'spotlight' with a 'central focus that may vary in size' (Posner et al., 1980, p. 171) or as a 'zoom lens' with an additional trade-off between processing width and processing depth (Eriksen & St. James, 1986), or gradually decreasing from the focus ('gradient models') (Downing, 1988; LaBerge & Brown, 1989) seems to depend on the task and is subject to ongoing discussion (Crundall, Underwood, & Chapman, 2002; M. W. Eysenck & Keane, 2005; Müller & Krummenacher, 2002).

Situations and tasks can be differentiated according to how much of the following three types of attention they require (Kluwe, 2006):

- selective attention;
- divided attention; and
- sustained attention or vigilance.

In the case where a situation requires successive processing and reaction to single relevant stimuli in the presence of several irrelevant stimuli, the term 'selective attention' is used. Trick, Enns, Mills & Vavrik (2004) further subdivided selective attention in the context of car-driving into four types, depending on whether it is exogenous or endogenous and whether it is automatic or controlled. Divided attention is used when there are several relevant stimuli which need processing in parallel. Divided attention is usually relevant in dual-task situations (see chapter 2.3.6). Finally, vigilance, or sustained attention, is needed for situations which require attention to be maintained for longer periods of time. In the driving context it can be defined as 'sustaining preparedness to respond to occasional events' (Groeger, 2000, p. 58).

Attention can be drawn to a stimulus voluntarily (top-down) or involuntarily due to stimulus characteristics (bottom-up). Top-down control is closely associated with the construct of schemata as used by Neisser (1976) (see Figure 13). When attention is drawn to a stimulus without voluntary control, the term orienting reflex (Rohrbaugh, 1984) or attentional capture (Kramer & McCarley, 2003; Simons, 2000) is used.

In a laboratory setting, Peterson et al. (2001) showed that 'implicit top-down effects … appeared capable of overriding even very powerful stimulus-driven processes in the control of attention' (cited in Kramer & McCarley, 2003, p.41). Similar effects were found for pre-cues which direct attention to subsequent relevant locations

(Luiga & Bachmann, 2007). Contrary to this finding, Theeuwes (2004) found that top-down processes are not capable of overriding attentional capture. It should be noted that these studies used highly artificial tasks in a laboratory setting and therefore might not be directly transferable to driving behaviour in a natural environment.

The respective importance of the two processes is assumed to be directly dependent on the characteristics of the stimuli and the environment used (for a review see Kastner & Ungerleider, 2000). The amount of attention that is under voluntary control has 'self-paced' properties similar to the regulation of speed. This means that drivers increase or reduce the amount of attention they give to a particular traffic situation depending on the perceived situational characteristics (Wikman, Nieminen, & Summala, 1998). Although the role of psychological factors in accident causation is always difficult to estimate, inattention or distraction is usually regarded as a highly relevant accident causation factor (Klauer, Dingus, Neale, & Sudweeks, 2006; Stutts et al., 2005). Aspects of attention which were assessed in laboratory tests have also been shown to be related to self-reported accident involvement (Arthur, Strong, & Williamson, 1994; Avolio, Kroeck, & Panek, 1985; Ball & Owsley, 1991; Myers, Ball, & Kalina, 2000).

The implications for rural road safety are thus clear: the situation has to provide drivers with correct information on how many resources are needed in this situation. This information must be presented in a way that matches expectations (top-down) and, depending on its importance, in a way that catches the driver's attention (bottom-up).

2.4.4 The Direct Approach to Perception

In contrast to Rumar (see chapter 2.4.1), Gibson (1986) stresses the bottom-up nature of perception. According to Gibson, the inherent physical properties of objects are directly acquired and determine human perception. These properties are called affordances. Affordances convey a meaning to the onlooker in the sense of being … –able (e.g. climbable) (for a summary of Gibson's theory of affordances see Jones, 2003).

Despite the differences between Rumar and Gibson, they both agree that perception is an active process. While Rumar stresses the importance of cognitive factors, Gibson sees movement as the crucial aspect in information acquisition. Movement of the body and the eye help to perceive the properties of objects and environments. Therefore, the human body as a whole becomes the organ of perception and not the eye alone. Through movement, the driver perceives information about depth, distance or speed. This information is perceived directly from the rate of change in the texture or the so-called optic flow field. The optic flow field can be seen as a cluster of vectors originating from changes in light due to movement. The focus of the flow field specifies the direction in which the observer is heading. Warren et al. (1991) showed that circular heading is also derived from the optic flow field when negotiating a curve.

Gibson (1986) further claimed that behaviour is guided by invariant information contained in the optic array. He found, for example, that objects of the same height are divided by the horizon in equal ratios no matter what the distance of the observer is (thus the name 'invariant'). Another example of such an invariant is tau (Tresilian, 1999). Mathematically, tau is equivalent to time-to-collision (TTC). Given that both ego speed and the speed of an oncoming object are constant, TTC can be calculated as:

$$TTC = \frac{d(t)}{v_i(t) - v_{i-1}(t)} = \frac{\text{(perceived) absolute object distance}}{\text{(perceived) absolute approch speed}} \tag{1}$$

The direct approach to perception assumes that TTC is not consciously calculated ('visual information about distance and velocity as such are not necessary …' Lee, 1976, p. 440), but directly perceived from the information available on the driver's retina. If the (monocular) retinal image of an object ahead of the driver increases, the driver approaches the object; if the retinal image decreases, the driver falls further behind of the object (or the car ahead moves away from the ego vehicle). In case of approaching, tau specifies how long it takes to reach the car ahead equal to TTC. Tau can be expressed in terms of the retinal image itself and thus equals the inverse of the rate of dilation of the retinal image of the obstacle (all formulas according to Lee, 1976):

$$\tau(t) = \frac{1}{\text{(rate of dilation of the retinal image of the obstacle)}} \tag{2}$$

Alternatively it can be expressed in terms of the visual angle subtended by this retinal image (θ) and its change over time which results in:

$$\tau(t) = \frac{\text{(angular seperation of any two image points of the obstacle)}}{\text{(rate of seperation of the image points)}} \tag{3}$$

$$= \frac{\theta}{(d\theta / dt)}$$

However, tau itself is not enough to describe the regulation of speed in distance keeping. This is because tau (or TTC) is defined assuming constant speed and thus does not take into account the change of speed over time. Therefore, a second variable was introduced by Lee (1976). This second variable, tau dot, is the derivation of tau over time and equals the rate of change in tau or TTC. Thus, it corresponds to the temporal change in acceleration or deceleration (Yilmaz & Warren, 1995). According to Lee (1976), tau dot is calculated as (see also Appendix A 1.1):

$$\dot{\tau} = -\left(1 + \frac{d \times acc}{v^2}\right) = -\left(1 - \frac{d \times dec}{v^2}\right) \tag{4}$$

where:

d = distance [m]

acc = acceleration [m/s^2]

dec = deceleration [m/s^2]

v = velocity [m/s].

A tau dot value of less than -1 (e.g. -2) corresponds to acceleration and a tau dot value equalling -1 corresponds to neither acceleration nor deceleration. In the latter case, a collision will take place at the value specified by tau. For tau dot values between -1 and -0.5 the driver will still collide with the vehicle in front as deceleration is insufficient to avoid collision. For tau dot values between -0.5 and 0 the car will come to a stop precisely before touching the vehicle in front. Only a tau dot value of exactly -0.5 results in constant deceleration (Yilmaz & Warren, 1995) (overview in Bruce et al., 1996; Groeger, 2002). While Lee (1976) assumed that drivers use a constant tau dot strategy, alternative tau dot strategies were also proposed and tested (Fajen, 2005; Rock, Harris, & Yates, 2006; Yilmaz & Warren, 1995).

Other authors questioned entirely a tau strategy and proposed alternative parameters (Tresilian, 1999). For the application to traffic, the difference between calculated TTC and directly perceived tau are of minor importance. Similar to other driving behaviour variables like time headway or speed, the preferred TTC values differ between driver groups. One variable which influences preferred TTC values was found to be driving experience (Cavallo & Laurent, 1988). Of similar importance for road safety is the distinction between 'local' and 'global' tau (Tresilian, 1991). While this distinction was originally made to reflect differences in calculating tau, it is also applicable when using calculated TTC: estimated TTC to a preceding vehicle decreases when the environment ('global') is enriched with additional elements.

Similar to TTC, Godthelp et al. (1984) proposed a time-based control mechanism for lane-keeping, the time-to-line or time-to-lane crossing (TLC). This is the time needed until either the left or right lane boundary is reached given that the present speed and heading are not changed. Thus, TLC is calculated similarly to TTC. Godthelp (1988) could show in occlusion experiments that drivers use a fairly constant (or 'invariant') TLC for corrective steering actions which is independent of speed level. Similarly, van Winsum and Godthelp (1996) assumed that the amount of speed reduction before curves is based on TLC. Salvucci (2006) developed a model for steering which is also based on changes of the visual angle with time. Thus, for rural road safety, the judgment of time remaining until an object such as another car, building or the lane boundary is reached by the driver's vehicle is an important behavioural variable.

However, despite the supporting evidence named above, drivers seem either not to totally rely on constant time-margins or they are far from perfect in correctly perceiving these time margins (Cavallo & Cohen, 2001; Gray & Regan, 1999; Groeger, 2002; Hesketh & Godley, 2002; E. R. Hoffmann & Mortimer, 1994; Sidaway, Fairweather, & Sekiya, 1996). Therefore, the usefulness of time margins is limited to short-term corrective driving behaviour rather than to anticipatory, open-loop behaviour for which expectations are more important.

2.4.5 The Perception of Ego-Speed

If drivers used a tau and tau dot strategy for driving as explained in the previous chapter they would not need to directly perceive speed. On the other hand, the application of such a strategy is restricted to closed-loop control of specific tasks such as approaching an object, car following or lane keeping. Situations or tasks which do not allow the use of tau and tau dot strategies still require the perception of ego-speed. This is especially the case for two situations.

Firstly, it is necessary for open-loop control before oncoming situations. Open-loop control requires that the expected appropriate speed for the oncoming situation is related to the current speed in order to infer required acceleration or deceleration. The determination of appropriate speed for an oncoming situation was described in chapter 2.3.8 where target variables such as subjective expected workload or risk were discussed.

The second situation where perception of ego-speed is necessary is driving on straight road sections. Here, a tau-like strategy could only be used if objects were situated directly ahead of the driver on the road; otherwise TTC or tau would be infinite as heading is not directed to objects on the side of the road. Speed regulation on straight road sections first of all requires that the legal speed limit is either kept or not exceeded above a self-set limit by the driver. As perception is context-dependent and yet objective information is needed for this task, this information can only be provided by the speedometer. In fact, by analysing fixations during driving, Schweigert (2003) found that drivers do use speedometers, especially on rural roads. On the other hand, the large variation in cross-sections of rural roads (FGSV 1995;

Matena et al., 2006) suggests that additional adjustments are required by drivers beyond simply keeping speed below the legal speed limit.

In both situations, ego-speed is perceived by the optic flow field as described by Gibson (1986). This perceived speed in relation to perceived demand or other motivational variables determines the adjustments made by drivers. In order to be perceived, the optic flow requires a structured environment which changes with movement. When driving on straight roads with no other traffic, the road itself does not offer much structure in its centre as this does not change with movement. Since drivers when driving on straight roads direct their fixations mainly to the centre of the road with exponential decline to the sides (Land & Lee, 1994), peripheral vision has to play a role in the perception of ego speed when defined by the optic flow. This is supported by Cavallo & Cohen (2001) who summarised several studies which stress the role of peripheral vision for speed perception (see also chapter 2.4.1). Peripheral vision also plays a role in lane keeping, but it has to be noted that performance decreases with increasing foveal eccentricity of the available information (Summala, 1998; Summala, Nieminen, & Punto, 1996; Warren & Kurz, 1992). A similar degrading effect was found by Tynan & Sekuler (1982) in the perception of speed: the speed of stimuli perceived peripherally is underestimated in comparison to when they are presented foveally.

Due to conflicting evidence the share of peripheral vision in the perception of ego speed cannot be determined at this point. Independent of peripheral or foveal emphasis, the perception of ego speed is facilitated by texture in the environment and the road itself where most fixations are directed. In the latter case, road markings could be used to provide additional texture to the environment.

3 A Psychological Model for Driving on Rural Roads

Today, several existing models and theories of driver and driving behaviour exist in parallel; however, this work requires a single framework which allows the deduction of hypotheses. Such a framework is provided by the model depicted in Figure 14, which is subsequently explained. As it is built on the theories introduced in the previous chapters, the relevant chapters are cross-referenced for details on specific elements in the model.

The starting point of the model is the objective road geometry and situation ahead. Both can be described by physical parameters such as radius, road width, elevation, sight distance and friction. The influence of the objective road situation on driving behaviour as well as on accident occurrence and accident severity is well documented (for an overview see Elvik & Vaa, 2004; Lamm et al., 1999). However, when it comes to exactly predicting the influence of geometry on behaviour or accident parameters, the models developed so far show a high variation (see chapters 2.3.9.2 and 2.3.9.3). One reason for this uncertainty in prediction other than methodological aspects is that driver and driving behaviour is not directly influenced by the objective road situation but rather by the perceived (and thus subjective) road situation.

However, this subjectively perceived road situation does not only depend on the geometry of the road. Of similar importance are other situational factors that are not as easily described as the geometry. These factors include singular cues which are part of the situation. Cues convey a message to the driver which can be sufficient to regulate behaviour. A road sign, for example, can almost be called the 'archetype of the discriminative stimulus' (Fuller & Santos, 2002, p. 49) and primes the relevant behaviour (Koyuncu & Amado, 2008). However, single cues can only be effective if their message is not concealed by the effect of 'filters' (see chapter 2.4.1).

In order to diminish the effect of filters, the properties of the situation should guide attention towards the singular cues; and the characteristics of the entire situation should be in accordance with the message conveyed by singular cues (Goldenbeld & van Schagen, 2007). However, the situation itself can also convey a message to the driver and thus serve as an 'integrated' (Fuller, 1984) discriminative stimulus itself without singular cues. A proxy assessment of the perception of a situation and its behavioural relevance can be done via subjective ratings such as the Road Environment Construct List (see chapter 4.2.2). The idea of the behavioural relevance of entire situations is conveyed by the concept of affordances which was laid out by Gibson in his theory of direct perception (see chapter 2.4.4).

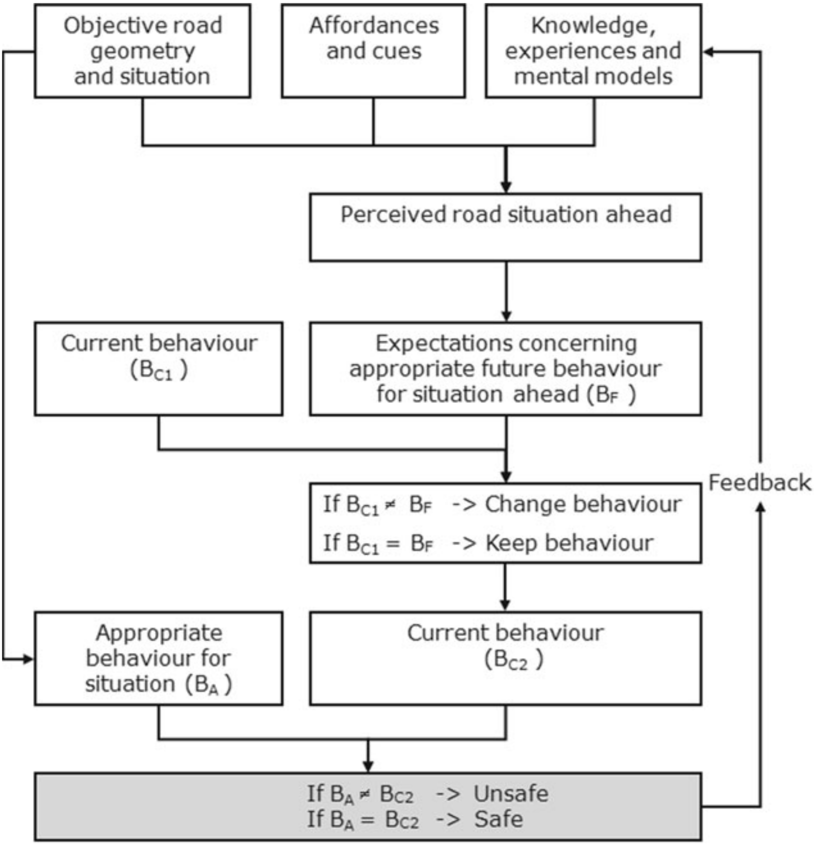


Figure 14. Driver and driving behaviour model for rural roads.

The perception of the objective road geometry and situation together with the cues and affordances are further guided by top-down processes. The relevance of such processes was first described by the Gestalt approach to perception (see Goldstein, 2002, for a summary) and is seen as a vital part of driving behaviour (Rumar, 1985). These top-down processes are guided by knowledge and experiences (both long- and short-term) which in their entirety form a mental model (see chapter 2.4.2).

The driver forms expectations about appropriate behaviour for the situation ahead based on the perceived road situation ahead and the past situation. These expectations might be relevant accident causation factors themselves as was shown by Steyvers (1993) and Petermann, Weller & Schlag (2007).

The expected appropriate behaviour for the situation ahead (B_F) is then compared to the current behaviour in the present situation (B_{C1}). If maintaining the current behaviour were to result in a discrepancy to the expected (or anticipated) appropriate behaviour, current behaviour is adjusted so that it will match the expected appropriate behaviour when the situation ahead is encountered.

Which behaviour is regarded as appropriate for the oncoming situation by the individual driver (open-loop control, see chapter 2.3.3) is not objectively defined but reflects individual preferences and might differ between drivers. The psychological mechanisms which are used to determine (consciously or unconsciously) an appropriate future behaviour are presumed to be the same as those used in common motivational models of driver and driving behaviour (see chapter 2.3).

Finally, the model predicts that if the resulting behaviour of this comparison process, now termed current behaviour B_{C2}, deviates from the objective appropriate behaviour for the situation, termed B_A, the behaviour in this situation is potentially unsafe.

Several assumptions are made in the driver and driving behaviour model for rural roads outlined in Figure 14. Some of these were tested in the subsequent empirical part of this thesis.

4 Empirical Validation

4.1 Overview and General Course of Events

In order to examine the propositions made in the driver and driving behaviour model for rural roads (see previous chapter, Figure 14), several studies were conducted. Of these studies, three (quasi-) experimental studies will be reported in this thesis:

- a laboratory study;
- a driving-simulator study; and
- a field study on rural roads conducted with an equipped vehicle.

These three studies served specific purposes with respect to the model-propositions.

The laboratory study was conducted to examine the role of the perceived characteristics of the road situation ahead on behaviour. The materials used in the study were pictures showing rural roads with manifold characteristics. To assess how drivers perceive these different characteristics, the Road Environment Construct List (RECL) (Steyvers, 1993, 1998; Steyvers et al., 1994) was used. Two questions had to be answered:

- Can the RECL factor structure be replicated in the German version and with the roads used in the laboratory study?
- Do the RECL items as proxy variables for the perceived road situation ahead allow the prediction of behaviour?

The latter question is thus directly related to the model. In order to answer these questions, the development of the original RECL is described together with its German translation and the statistical background of factor analysis. An additional aspect concerns individual differences and thus a potential broadening of the original model. Whether individual factors must be emphasised in the model was prototypically tested with structural equation models.

The driving-simulator study was conducted to examine the role of specific cues and affordances which are an integral part of the driver and driving behaviour model for rural roads. Their effect on both perception of road characteristics and on behaviour was examined. It constitutes an extension of the preceding laboratory study because it allows the experimental variation of single cues and other elements. Furthermore, behaviour is directly measured instead of collected as ratings.

Additionally, the results of the laboratory study can be applied in a more realistic driving context with respect to relevant aspects of the perception of road characteristics.

The field study was mainly conducted to analyse the last part of the model, that is, the effect of behaviour on accidents. According to the final part of the model, a situation is unsafe if actual behaviour deviates from appropriate behaviour. The focus in the field study thus shifts from open-loop control as in the laboratory and simulator study to closed-loop control. The research paradigm chosen was the comparison of high accident-rate curves with low accident-rate curves of similar geometry. Behaviour in the low accident-rate curves served as a reference of appropriate behaviour. In addition to speed, behaviour was also measured by recording gaze-behaviour and reaction time to a peripheral detection task.

In addition to these studies, further studies were conducted during the project. Some of these re-analysed data provided by Dr. Wagner and Prof. Richter (Chair of Work and Organizational Psychology at TU Dresden). Information on the original study and its results can be found in Wagner (2000), Richter et al. (1998), Richter et al. (1996), and Heger & Weise (1996). Although the re-analyses of speed in combination with her own data provided some interesting results (Petermann, 2006), the re-analyses of the psycho-physiological data with respect to singular geometric elements was less promising (Wendsche, Uhmann, & Meier, 2006). Amongst other factors the latter is because psycho-physiological variables usually require some time until they show an effect. This makes it difficult to relate them to single geometric elements such as curves. Selected results of these studies are described in Weller & Schlag (2007) or have already been published elsewhere (Petermann et al., 2007). All of these studies served as valuable input for this thesis.

The statistical analysis of the data was conducted using SPSS. The preparation of the data was conducted using either SPSS syntax commands (Brosius, 2005) or a combination of VBA-scripts and SQL-commands in Microsoft Access. In addition, the programmes RoadView (Dietze, 2007) and WatchOut (Schulz, 2007) were used which were both tailor-made at the Chair of Road Planning and Road Design at TU Dresden and which are described in the relevant chapters.

4.2 The Laboratory Study: The Role of Perceived Road Characteristics

4.2.1 Introduction and Rationale of the Laboratory Study

The laboratory study served several purposes. Its overall goal was to shed light on the role of perception on driving behaviour. In order to conduct this task, a method first had to be found which allowed the assessment of ‘perception’. Such a method has already been developed by colleagues in the Netherlands with the Road Environment Construct List, in short RECL (Steyvers, 1993, 1998; Steyvers et al., 1994). The RECL had to be translated into German before it could be applied in the laboratory study. A prerequisite for the further application of this German version of the RECL was a stable factor structure which additionally replicates the factor structure which was established in the original RECL. These methodological issues constitute the first part of the results for the laboratory study.

After these methodological issues, a first attempt was made to test the model assumption that the perception of the characteristics of a road situation as collected with the RECL items allows a prediction of behaviour (see the driver and driving behaviour model for rural roads in Figure 14). In the laboratory study behaviour was not measured directly but was collected with ratings concerning the speed participants regarded as appropriate. Despite this limitation, this analysis is a worthwhile contribution to the model validation. This is because of the advantages provided by a controlled experimental environment in the laboratory in combination with the large variety of roads which could be tested with this experimental setup.

A third part of the results derived from the laboratory study dealt with the role of individual factors on both perception and behaviour. Exploratory analyses dealing with this issue were conducted and mainly served the purpose of determining whether the role of individual factors which were so far only indirectly included in the model ('knowledge and experience') needed to be stressed further.

Finally, the laboratory study was conducted to contribute to a self-explaining rural road design in Europe. The results of this task have already been published in Weller, Schlag, Friedel & Rammin (2008) and are only referred to where necessary for the context of this thesis.

4.2.2 The Original Road Environment Construct List (RECL)

Four Dutch studies (Kaptein, Janssen, & Claessens, 2002; Riemersma, 1988; Steyvers, 1993, 1998) served as input for the development of the Road Environment Construct List (RECL). The RECL was chosen for different studies in this thesis because of several reasons which will be summarised at the end of this chapter. Before this, its development will be described in more detail.

The RECL was developed by Steyvers in order to analyse an existing safety problem on two roads in the Netherlands. Both roads had a higher rate of single vehicle accidents than comparable roads. In addition, both roads differed from one another with respect to these accidents. As both roads were geometrically identical, this latter difference could not be explained by road geometry. However, the roads differed with respect to the environment in which they were located: one road was in a more or less barren environment, a so-called polder (i.e. reclaimed land), the other environment was enriched with vegetation. The underlying assumption when developing the RECL was that the perception of these environmental differences could explain the differences in accident figures. The theoretical assumptions are comparable to the ones formulated in the model depicted in Figure 14: drivers are expected to construe a road based on their perception of the environment. This subjective road representing a ‘personal construct’ (Kelly, 1955/1991) is in turn expected to influence behaviour.

The development of the RECL was conducted in three steps (Steyvers, 1993):

- construct (or item) elicitation;
- construct selection; and
- construct rating.

Slides of the different locations of the two roads referred to above (three locations for each road) under different traffic (traffic vs. no traffic) and light conditions (daylight vs. dusk) were used to collect the RECL items. The resulting 24 slides of the two road sections were freely described by 24 participants. These descriptions were condensed to 26 constructs by psychologists.

In the next step, these constructs were used by another group of participants to rate the 24 slides. This time, the slides were presented pair-wise with the pairs matched for traffic and light condition (resulting in 12 pairs). Participants were asked to select one out of the 26 constructs which best described the differences between the two slides and to name a corresponding antonym. The resulting construct-contrast pairs were then used to rate single slides. As different participants named different contrasts, each contrast and construct had to be named at least ten times to be selected for this step. The resulting constructs were applied as a five point rating scale. Where a construct was disproportionately rated ‘3’ (meaning ‘I do not know’), it was discarded.

This resulted in 18 constructs which were finally presented as items on a six-point rating scale to another group of 84 participants. The resulting data was averaged for the three road locations and the remaining data was arranged in a matrix with the 18 constructs as columns and the 84 participant times conditions (2 × 2 × 2) as rows. Steyvers justified this transformation of the dependent data structure into an independent one with the high test/retest reliability he found in his data. This was between .80 and .90 when the ratings were repeated after a break by the same participants (Steyvers, 1993) and was of similar height when the same road sections in the different traffic and light conditions (see below) were presented to different participants either as slides or as films (Steyvers et al., 1994). Arranging the data in the way introduced above is also described in Backhaus, Erichson, Plinke, & Weiber (2006, pp. 326).

This matrix was used as input data for a principal component analysis (PCA) with Varimax rotation. The results of this 18-item PCA are reported in Steyvers (1993). Some minor changes were made as a result of the findings reported there resulting in the final 16-item solution which is reported in Steyvers et al. (1994) (see also Appendix A 2.2). Two items (oppressive and safe) were deleted and ‘increases attention’ and ‘increases concentration’ were renamed (‘lowers’ instead of ‘increases’) in order to increase the variation in the construct direction. This final version was presented to a new group of 64 participants who rated the same roads in the same environment and traffic conditions. However, this time films were used instead of slides (Steyvers et al., 1994). Data was arranged in the same way as above and again PCA with Varimax rotation was performed. This resulted in three factors which accounted for 66.8% of the variation in the data (Steyvers et al., 1994). The three factors were termed ‘Hedonic Value’, ‘Activational Value’ and ‘Perceptual Variation’.

The term ‘Hedonic Value’ needs some further explanation. Hedonic value is derived from the work of Berlyne, ‘a term meant provisionally to cover both reward value, as judged by the capacity to reinforce an instrumental response, and preference or pleasure, which is reflected in verbal evaluations’ (Berlyne, 1970, p. 284). The verbal evaluations referred to in this citation were collected by Berlyne with a single rating of ‘pleasing’ on a seven point verbal scale (Berlyne, 1970). The objects which were rated consisted mainly of artificial, drawn, two-dimensional symbols of different complexity (examples can be found in Berlyne, 1972). In the context of this thesis it is interesting to note that Berlyne found evidence that this rating of pleasantness depends not only on the characteristics of the object but differs also depending on the activation or arousal of the participant who rates the objects (the terms activation or arousal are used interchangeably by Berlyne) (a summary of these results is found in Machotka, 1980). These findings are interpreted by Berlyne with respect to the Wundt curve which links pleasure (positive and negative 'Werthe der Lust' Wundt, 1874, p. 432) to stimulus intensity. Both of these influenced Zuckerman (1994; 2007) who found that high sensation seekers preferred more complex figures (see also chapter 2.2).

The variety of items constituting the factor ‘Hedonic Value’ and the other two factors reflect the complexity of this theoretical background:

- Hedonic Value: threatening, dangerous, gives a good view (-1)[7], demanding, spacious (-1), enjoyable (-1), irritating, relaxing (-1);
- Activational Value: lowers attention, lowers concentration, increases wakefulness (-1), increases alertness (-1);
- Perceptual Variation: changeable (-1), monotonous, boring.

---

[7] (-1) indicating negative factor loadings of the item on the factor.

Initially it would appear that the naming of the factors is misleading when the polarity of the items in the factors is taken into account. For example, one would assume that the item 'changeable' loads positive on a factor named 'perceptual variation'. However, from the figures in the publications and the somewhat rudimentary description of the recoding process ('*Where necessary* ratings for negative simple weights of *some* construct scores were mirror-reversed …' Steyvers et al., 1994, p. 505, own accentuation), it can be inferred that for the interpretation of the results as presented below, the inherent meaning of the factor names was used. This is supported by the fact that instead of the factor scores derived from the factor analysis, the averaged factor-wise rating scores were used for the analysis of differences between the different conditions (light, traffic, roads, see above) (Steyvers, 1993).

Several ANOVAS were calculated with these 'factor values'. The most important result with respect to traffic safety can be seen in the significant main effects between the two roads. The Polder road, the road with the higher accident-rate, showed in the daylight, no-traffic condition

- less Activational Value;
- less Perceptual Variation; and
- higher Hedonic Value.

Further on, the ratings for the road with the higher accident-rate were less stable and were far more influenced by traffic and light conditions than the low accident-rate road. The reader is referred to the original work for more detailed results in particular regarding the different traffic and light conditions.

The reasons for using the RECL can thus be summarised as follows:

- Throughout the laborious item elicitation process, an exhaustive number of aspects were collected along which rural roads can be distinguished.
- These aspects were carefully selected and condensed to a quite small number of marker items, again in a way to meet quality standards in psychological research.
- The final selected items and the factors found comprise aspects of all the theories regarded as relevant to driving on rural roads (see methodology), namely aspects of demand, risk and aspects of visual perception.
- The factors found showed a high reliability and could successfully distinguish between high and low accident-rate rural roads.
- With the emphasis on perceived road situation ahead, the factor values allow testing of several aspects related to feed-forward, open-loop strategies in driving.

In order to be applicable in a German laboratory study, the RECL was translated into German.

4.2.3 The German Version of the RECL

The 16 RECL items which were available in Dutch and English (Steyvers et al., Steyvers, 1993; 1994) were translated into German and tested for comprehension and applicability with small samples in preliminary tests (details in Petermann, 2006; Voigt, 2007). As already indicated, RECL items were used in different studies in this thesis. The versions used differed in minor aspects from one another which will be described in the respective chapters. In addition to these minor differences, another aspect became increasingly important during the various studies performed for the thesis. This aspect concerned the factor structure of the RECL. It was initially assumed that the original factor structure could be replicated owing to the high quality of the factor structure reported in Steyvers (1993, 1994). However, in almost all of the student papers and diploma theses which contributed to this thesis (Friedel, 2005; Petermann, 2006; Rammin, 2006; Voigt, 2007) the authors found differences in the factor structure. While again details are reported in the respective chapters where necessary, potential reasons for these differences must be discussed here. This is considered necessary as it contributes to the decision of whether factor scores or item values can (or must) be used for the subsequent analysis.

There are several potential explanations for the deviations between different factor solutions. The most important aspect is seen in the differences of the stimuli used here versus the stimuli used by Steyvers. Steyvers merely used two different straight road sections, albeit in different traffic and light conditions. In the studies presented here, a higher variation of rural roads were used, partly also including curved sections. A further aspect concerns the translation of items into German. Their equivalence to the original Dutch items and their English translation could not be assessed, although the translated items were preliminarily tested for their applicability and although of course care was given to translate the items as closely to the original meaning as possible.

The obvious low reliability of the factor structure between different studies questions the appropriateness of using factor solutions for further analyses. However, each of the different factor solutions proved successful in their own right in the studies mentioned above (see also Petermann et al., 2007; Weller, Schlag et al., 2008). In order to judge the quality of these factor solutions and the appropriateness of using factor solutions at all, the statistical background behind factor analysis has to be understood. Therefore, an extra chapter was included in this thesis which exclusively deals with factor analysis (chapter 4.2.4.6).

4.2.4 Methodology

4.2.4.1 The Stimuli

The stimuli used in the laboratory study were pictures of rural roads which were presented to the participants on a computer screen (see chapter 4.2.4.3). The pictures comprised a large variety of rural roads with an equally broad variety of different characteristics. All pictures showed roads in Saxony and Brandenburg and were taken in summer during dry weather conditions from the perspective of a car driver both in height and in position on the road. No other road users were visible in the pictures. For additional analyses, some pictures were edited on a computer so that the edited picture differed from the original picture in just one element. In total, 25 pictures were thus collected. After a preliminary study reported in Friedel (2005), Rammin (2006), and Weller et al. (2008), 21 of these 25 pictures were selected and rated by participants. All pictures are shown in Appendix A 2.1.

4.2.4.2 The Sample

The sample consisted of 46 participants (21 female and 25 male) aged between 20 and 65 years (M = 40; SD = 5 years). All participants had held a valid driving licence for between 2 and 47 years (M = 20; SD = 14 years). The average annual distance driven for the last two years varied between 500 and 80,000 km (M = 21,152 km, SD = 17,447 km).

4.2.4.3 General Course of Events

In terms of research methodology, the study was conducted along similar lines to the aforementioned Dutch studies from Kaptein et al. (2002), Riemersma (1988), and Steyvers (1993; 1998). The pictures were presented on a computer screen (19’’) to the participants who were asked to fill out the questionnaire (paper & pencil; no time constraints). The order of presentation of the pictures was balanced, so that the first picture of one experimental run was the last picture for the next participant. The questionnaire consisted of the translated 16 items of the RECL as described in chapter 4.2.3 (see also Appendix A 2.2). In addition, one item asked participants about the subjective appropriate speed in the situation seen in the picture (rated in km/h). All RECL items were rated on a six-point Likert scale with verbal anchors. The questionnaire consisted of further items, which were described in detail in Friedel (2005) and Rammin (2006) and which were not analysed for this thesis.

4.2.4.4 Introductory Remarks to Statistical Testing

This chapter introduces some statistical concepts which are used throughout this thesis. Although the concepts introduced are usually well known by experimental researchers, it is regarded as a helpful introduction to the chapters on the selection of specific statistical methods.

Prior to statistical testing the researcher's assumption with regard to the effect of a treatment must be formulated as a hypothesis. In general, two types of hypotheses are differentiated. The alternative hypothesis assumes that a treatment results in an effect whereas the **null hypothesis** assumes the absence of this effect.

The assumptions formulated in the hypotheses are the models which are tested statistically. Statistical testing as used here means to quantify the probability that the data found in the sample is in accordance with the hypothesis. Following the principle of falsification laid out in the critical rationalism by Popper (1934/1989, cited in Bortz & Döring, 2006) only the falsification of the null hypothesis can be tested. Therefore, error-probabilities are given as a result of statistical tests. The error probabilities indicate the probability of alpha- or Type I errors. An alpha-error is the probability that the null hypothesis (H_0) is wrongfully rejected because of the results found in the sample despite the fact it is valid in the population (Bortz, 2005).

Now the natural question is where the error probabilities come from. A prerequisite to calculate the error probabilities is the assumption of a certain distribution of the values in the population. Important distributions are the chi-square distribution, the t-distribution and the *F*-distribution, some of which are also normal distributions. By z-standardizing the x-values (see formula in Appendix A 1.2), every normal distribution can be transformed into a standard normal distribution which is characterised by a mean of zero and a standard deviation of one. For all distributions named so far, the graph of the distribution of the values is a probability density function for which the total integrated area below its graph is one. The area cut off the graph by a value found in the sample indicates the probability that this value could also be found in the population and thus denotes the error probability. The value which is used to describe the sample is the test statistic. Although its actual calculation differs depending on the distribution assumed it can usually be simplified according to Field (2009) as:

$$\text{test statistic} = \frac{\text{variance explained by the model}}{\text{variance not explained by the model}} = \frac{\text{effect}}{\text{error}}. \qquad (5)$$

The scientific community has agreed upon standards concerning the interpretation of error probabilities because they cannot by definition be used for the falsification of the null hypothesis (see Bortz & Döring, 2006 for a discussion). These standards are based on the work of Fisher (1925, cited in Bortz & Döring, 2006; Field, 2009) and are set at five or one percent error probability, below which the falsification of the null-hypothesis is accepted. Subsequently, the results are called significant if the error probability is below five percent or highly significant if it is below one percent. In both cases, the null hypothesis can be rejected and it is assumed that there is indeed an effect as predicted by the alternative hypothesis. If it is only important to know whether a result is significant or not, this can also be done by selecting critical values. The critical values are the values below and above which the test statistic would result in a significant result.

Two critical values have to be selected from the tabulated values because of the two directions of the alternative hypothesis. In order to jointly account for five percent each value has to cut off 2.5 percent of its side or 'tail' of the distribution. Therefore, such hypotheses are called two-tailed and the respective significance is the two-tailed significance. If supported by evidence from existing results or theories, one could also formulate a directional hypothesis. In this case, there is only a single critical value which cuts off five percent of the distribution at one tail of the distribution.

Due to the difference between one- and two-tailed testing, effects which are tested one-tailed can be smaller to get significant than effects which are tested two-tailed. In SPSS, the default results in the output are all given for the two-tailed significance. Consequently, the results reported in this thesis are also the two-tailed results, despite the fact that sometimes directional hypotheses were tested (see chapter 4.3.2). Such conservative testing was used in order to account for the rather weak evidence for one-tailed testing due to the relatively new approaches used. However, in the case of t-tests, the two-tailed error probabilities can simply be halved to result in the error-probabilities for the one-tailed tests, if required (Field, 2009). This is not possible for the ANOVAs or other tests which test more than two conditions (see also Field, 2009).

4.2.4.5 Statistical Methods Used for the Research Questions

In chapter 4.1 three purposes of the laboratory study were named:

- validation and replication of the factor structure of the original RECL with the German version of the RECL;
- prediction of behaviour with these RECL factors; and
- analysis of the role of individual factors for both perception and behaviour.

The statistical methods which were used to analyse each of these tasks are reported separately for each task.

Independent of the task itself, the selection of a specific statistical method depends on whether the data within and across all experimental conditions is distributed normally or not (definition of normal distribution in Bortz, 2005). For this thesis, normal distribution was tested with the Kolmogorov-Smirnov test (K-S test). The K-S test tests the null hypothesis that there are no differences between a normal distribution and the distribution found in the sample. In order to assume the validity of this null hypothesis (strictly speaking: to not falsify it; see preceding chapter) the error probabilities should be very high, that is, the test should not be significant. In cases where the K-S test indicated deviations from the assumption of normality, the histograms of the data were also inspected visually. This additional inspection was done because the K-S test can become significant even with minor deviations from the assumption of normality, especially with large sample sizes (Field, 2009).

The first task (replication and validation of the RECL factor structure) basically required the application of factor analysis. The statistical background of factor analysis is described in detail in the next chapter. The decision for this detailed discussion of the method was made because of several reasons: firstly, outcomes of the factor analysis (the factor structure and the factor values) were regarded as central for the model and were also planned to serve as input for the other two tasks. Secondly, factor analysis is the basis for other statistical methods. For example, understanding the background of factor analysis is also useful for understanding structural equation models which are used in their most basic form for the third task (see below). Thirdly, although factor analysis is widely applied, its statistical background is badly neglected. This is because of the seeming ease with which it can be applied and interpreted owing to modern statistical software packages and because of the exploratory nature in which it is often used (applied only 'when one has no other idea what to do with the data', remark of a professor of methodology at a congress in Nuremberg, 2006). However, the most important reason is that the preliminary analysis of the data already indicated that the factor structure deviated from the original factor structure. Understanding factor analysis in detail was seen as a prerequisite to judge the outcome of the analyses conducted for this thesis.

For the second task (prediction of behaviour), linear regression analyses were used. The exact method and the variables used for the analyses depended on the results of the first step and are therefore explained in the chapters in which the second task is discussed. The general idea behind linear regression is that the values of the dependent variable can be approximated by a function which includes the weighted values of the independent variables. This function defines the regression line and is determined with the method of least squares. These least squares in turn correspond to the sum of squares as described below in the paragraphs about the ANOVA.

In addition to the assumption of normally distributed data, linear regression analyses have some assumptions which should be met. As the name implies, the relationship between the predictor variable(s) and the dependent variable has to be linear. Despite some bell-shaped functions in psychology (e.g. between demand and performance, see chapter 2.3.5) this is assumed for the data here because presumably no extreme conditions were included in the stimuli. Linear regression further assumes that the residuals are independent which is tested with the Durbin-Watson statistic. This statistic should be between 1.5 and 2.5 with values below one or above three indicating a violation of this assumption. In addition, no multicollinearity is assumed. Multicollinearity means that the values of one predictor variable are a (perfect) combination of all other predictor variables, indicated by high correlations. It is tested by the VIF and tolerance statistics (tolerance is the reciprocal of the VIF statistic). Tolerance should be above 0.1 and the VIF value should be below 10 (all numerical limits taken from Brosius, 2008; Field, 2009).

Several parameters can be used for the interpretation of a regression. Firstly, the overall quality of a regression is indicated by the parameters *R* and *R* square. *R* is the (multiple) correlation coefficient between the predictors and the outcome of the regression function. Thus, *R* square gives the amount of variance of the dependent variable which is explained by the regression function. R square is adjusted to account for the number of the predictor variables when there are several predictor variables (Brosius, 2008). Secondly, an ANOVA is used to indicate whether the outcome of the regression function is significantly better suited to represent the data than simply using the mean of the dependent variable (Field, 2009).

In addition, the contribution of each independent variable in the regression is indicated by a *t*-test which tests whether the coefficient *B* is significantly different from zero. In a univariate regression, *B* is the slope of the regression line. In general, *B* indicates the amount of change in the dependent variable given the independent predictor variable is increased by one unit and all other independent variables are kept constant (Field, 2009). Because *B* depends on the units of the predictor variable, the standardised coefficient Beta β is also given which represents B in the situation that all variables were *z*-standardised before the regression (Brosius, 2008). The units for Beta are no longer the original units but are the standard deviation of the variable. Betas can easily be compared between variables because the standard deviation of a variable is independent of its unit. However, the interpretation is still the same: one unit change in the independent variable results in Beta times the unit changes of the dependent variable (Field, 2009).

Prior to applying regression analysis, several correlations between items were calculated (details are described in the respective chapter). Because data were not normally distributed in all cases, the non-parametric coefficient Kendall’s Tau was calculated in combination with Pearson’s *r* (see next chapter for the formula). Correlation coefficients are a measure of effect size (Field, 2009) and thus their interpretation is comparable (Cohen 1988, 1992, cited in Field, 2009):

- $r = .10$ is a small effect;
- $r = .30$ is a medium effect; and
- $r = .50$ is a large effect.

Brosius (2008) gives similar guidelines for the verbal interpretation of correlation coefficients. Again, squaring Pearson's r denotes the amount of explained variance between the two variables (Field, 2009).

For the third task (role of individual factors), several statistical methods were applied. In the first step, picture-wise multivariate analyses of variance (MANOVAs) were calculated with rated speed and the two marker items (to be defined later) as dependent variables and age group as factor. Only the youngest and the oldest age group were used for this exploratory analysis of differences. In SPSS, MANOVAs and ANOVAs for repeated measures are implemented as generalised linear models (GLM). GLMs combine the methods of simple ANOVAs with linear regression analyses (see above) and thus allow much more complex data analyses such as using covariates or several factors (an overview of further extensions is given in Brosius, 2008, the application of regression analysis in ANOVA is described in Field, 2009). Since the term ANOVA is more common, it is used instead of, or interchangeably with, GLM in this thesis.

The assumption of normality in a simple univariate ANOVA is extended in MANOVAs to all variables (multivariate normality). A further general assumption of ANOVAs is the assumption of the homogeneity (equality) of the variances in the different conditions. In ANOVAs these conditions are defined by the different conditions of the between-subjects factor (e.g. different age groups) and are tested with Levene's test. In MANOVAs, this test is conducted separately for every dependent variable. Levene's test tests the null hypothesis that the variances in the different groups are equal and therefore should not be significant. Levene's test is quite robust against violations of the precondition of normal distribution (Bortz, 2005, p. 286).

In addition to the assumption of homogenous variances between groups, MANOVAs assume the homogeneity of the covariances. The combined assumption of homogenous variances and covariances is tested with Box's test which tests the null hypothesis that the variance-covariance matrices are the same across groups. Where this test is significant, the assumption is not met. However, Box's test reacts progressively in cases of non-normally distributed data, that is, it indicates non-homogenous variances despite the fact that they are homogenous (Bortz, 2005, p. 619). Therefore, in equal group sizes, Field (2009) proposes completely disregarding the Box's test statistic. In any case, a robust algorithm of the MANOVA should be chosen.

Of the different algorithms used in the MANOVA part of the GLM, Pillai's trace is regarded as the most robust statistic (Bortz, 2005; Field, 2009; Rudolf & Müller, 2004; SPSS 16 Tutorial) and was therefore chosen to be reported in this thesis. However, in unequal sample sizes, Pillai's trace loses its robustness (Bortz, 2005; Field, 2009; SPSS 16 Tutorial). This is even more the case if sample sizes within the groups are less then ten (Rudolf & Müller, 2004, p. 80) and Box's test is significant (Field, 2009). Univariate statistics are given for each of the dependent variables in addition to the overall multivariate results. These univariate statistics are calculated within MANOVA by using univariate ANOVAs. As stated above, the assumption of homogenous variances between the age groups was tested separately for each of the dependent variables with Levene's test.

ANOVAs in general are based on a comparison between the variance between conditions and the variance within conditions. Calculating the variances is based on the sum of the squared differences referenced to the degrees of freedom. The relationship between the variances as given in the preceding chapter gives the *F*-statistic which can then be tested for significance (see Appendix A 1.6 for the calculation of the *F*-statistic).

Due to the fact that the significance is subject to sample size, a measure of effect size is given additionally for the MANOVA results. For (M)ANOVAs this measure of effect size is called eta squared (η^2). Eta squared is the total proportion of variance explained by the factor(s) (here: age group) and is in fact R squared (see the paragraph on the regression analysis). Its calculation is again given in Appendix A 1.6. By taking the square root of eta squared, the result is the correlation coefficient and effect size *r* which can be interpreted as given above (see the paragraph on correlations). As well as the effect for the MANOVA, an effect size is also given for the univariate part of the ANOVA. Here, the effect size is called partial eta squared (partial η^2) and is calculated as given in Appendix A 1.6. In contrast to eta squared, partial eta squared does not indicate the amount of overall explained variance but only the amount of variance which is not already explained by other variables in the equation (Field, 2009, p. 415). It is important to note that the effects cannot be added up for the interpretation of partial eta squared.

Differences between age groups were further included in repeated-measures ANOVAs. Details concerning the selection of age groups and the variables are given in the respective paragraph. In repeated-measures ANOVAs the central assumption of the homogeneity of variances is applied to the homogeneity of the variances of the differences between the different levels of the repeated-measures factor (Field, 2009). This assumption is called sphericity and is tested with Mauchly's test. Sphericity can be assumed when this test is not significant (Field, 2009; Rudolf & Müller, 2004; SPSS 16 Tutorial). Subsequently, the statistics for sphericity assumed in the SPSS output file can be used. As the calculation of a difference requires at least two variables and the comparison of the variance of these differences requires at least one additional variable, Mauchly's test is not defined for less then three variables (Bortz, 2005; Field, 2009). Although this was the case for the analyses calculated for the laboratory task, the consequences of a significant result of Mauchly's test are nevertheless described because they are needed later for the analyses of the simulator data.

Where sphericity cannot be assumed, the degrees of freedom used for the testing of significance can be corrected according to different algorithms (details are given in Bortz, 2005). Of these, the Greenhouse-Geisser is the most conservative, especially for small samples (SPSS 16 Tutorial) and is therefore given in this thesis in the case Mauchly's test is significant. In the SPSS-output file, MANOVA results are given in addition to the results of the repeated-measures ANOVA. While Kinnear & Gray (2009) suggest ignoring these results completely, Field (2009) suggests using them additionally where the assumption of sphericity was not met. For this thesis, the latter procedure was adopted. For the between-subjects factor age group which was included in addition to the repeated measures within-participants factor, the assumptions concerning between-subjects factors already discussed for the MANOVA had to be tested and met accordingly for the repeated-measures ANOVA.

Finally, structural equation models (SEM) were used to further analyse the effect of age on both perception and expected appropriate behaviour. This was done by using AMOS, a software package distributed with SPSS which allows the user-friendly modelling of SEM. The basic terms and ideas behind SEM are introduced in the next paragraphs which were written using the books of Airbuckle (1997), Backhaus et al. (2006), Rudolf & Müller (2004) and Byrne (2001).

A structural equation model consists of a few basic design elements which can be combined to result in quite complex models. Firstly, a distinction is made in SEM between exogenous variables which represent the independent variables and endogenous variables which are synonymous with the dependent variables. Both exogenous and endogenous variables usually consist of a combination of latent and manifest variables. Manifest variables represent the observed data, for example, the responses to an item in a questionnaire. In contrast to the manifest variables, the latent variables cannot be directly observed. Latent variables are essentially factors which are determined via factor analysis (see above and the next chapter) from the manifest variables. Strictly speaking and according to the rationale of factor analysis, the latent variable influences the manifest variable and not vice versa. Graphically, this relationship is depicted by an arrow pointing from the latent variable towards the manifest variable.

Statistically, this arrow – or path – represents the correlation between the latent and the manifest variable. The path coefficient is thus the correlation coefficient and is equal to the factor loading of an item in factor analysis. Squaring it denotes the amount of explained variance of the item by the latent variable or the factor. In addition, similar to the rationale of ANOVAs, the values of the manifest variables are also subject to a residual variance. This residual variance is present in all manifest variables and is also depicted graphically. As the total variance in standardised variables equals one, the residual variance is one minus the amount of explained variance.

In addition to the relationship between manifest and latent variables, the relationship between exogenous and endogenous variables must be determined. Graphically, this relationship is again depicted by an arrow pointing from the independent exogenous variable to the dependent endogenous variable. Statistically, it is determined by regression analysis. Thus, the standardised path coefficient equals the Beta-weight in linear regression analysis with the respective interpretation (see above).

Finally, also the latent variables can be inter-correlated (separately for endogenous and exogenous variables). Because this covariance, which is the non-standardised correlation, is not interpreted as a causal relationship, the path is graphically depicted as a double-headed arrow. Such a relationship might exist despite the fact that factor analysis methods such as the Varimax rotation strive for independent factors. In case of highly correlated factors the issue of multicollinearity might consequently be an issue.

SEMs were used in this thesis to test two alternative relationships between variables (details in chapter 4.2.5.7). For such tasks, AMOS provides a variety of indices which can be used to determine the quality of a structural equation model. While it is not possible to thoroughly discuss these goodness-of-fit indices in the context of this thesis, the following rules-of-thumb can be used for their interpretation (based on Arbuckle, 1997; Backhaus et al., 2006; Byrne, 2001). The relationship χ^2/df should be ≤ 2.5 and the p-value for χ^2 should not be below 0.1. The goodness-of-fit value (GFI) should be close to one and this also applies to the adjusted GFI (AGFI) which takes into account the degrees of freedom. Possibly the strongest statistic is the RMSEA value in combination with the probability of close fit (PCLOSE). RMSEA values of ≤ 0.05 indicate good fit, of ≤ 0.08 reasonable fit and of ≤ 0.10 unacceptable fit. PCLOSE tests the null hypothesis that RMSEA is indeed ≤ 0.05 and thus should result in an error probability PCLOSE which is not significant.

In addition to these goodness-of-fit indices, the standardised residual covariance matrix and the modification indices can be used to determine the quality of different models. The standardised residual covariance matrix is the difference between the covariance estimated by the proposed model and the covariance found in the data. These values should therefore be close to zero, whereas values > 2.58 (Byrne, 2001, p. 89) or > 0.1 (Backhaus et al., 2006, p. 383) indicate a bad model fit. Modification indices basically provide an estimation of the change in the χ^2-statistics if parameters which are fixed in the model were freely estimated. High modification indices indicate that the path between the two variables for which the modification index is given should also be included in the model.

The selection of an estimation method for the SEM statistics depends on the distribution of the data. Where data are normally distributed, maximum likelihood estimation can be used which results in the most accurate estimation (Backhaus et al., 2006). Due to the combination of factor analysis and regression-analysis methods, the respective assumptions also apply to SEM. Additional assumptions concern the sample size. In fact all estimation methods of SEM require sample sizes larger than 100 (see Backhaus et al., 2006, p. 371), which was not the case for the data in this thesis. Alternative methods to estimate the required sample size, also given in Backhaus, require using five times the number of parameters to be estimated, which was achieved (see chapter 4.2.5.7).

SEM strictly speaking further requires that where a proposed model is rejected, new data must be collected and the revised model has to be retested with the new data. Where the sample size is large enough, it is further useful to halve the sample size and use the second half as a validation sample for the structure found in the first half. For the data in this thesis, neither of these two approaches was possible due to the small sample sizes. Therefore, the analyses conducted in chapter 4.2.5.7 are regarded as preliminary input for future model adjustments rather than as statistical testing in its strictest sense.

4.2.4.6 Statistical Background: Factor Analysis

This chapter was included because of the reasons named in the preceding chapter. Two books were used as the main sources for this chapter: Bühner (2006) and Pett, Lackey, and Sullivan (2003). Additional sources are given in the text.

The basic idea behind factor analysis methods is that the measured values of a large number of items can be aggregated into a comparatively small number of latent factors. These latent factors represent constructs which otherwise cannot be measured. The constructs incorporate a large variety of different, yet related aspects which are represented by the items. Only measuring one item would give an incomplete picture of the real construct value. Furthermore, the number of variables can be reduced considerably by using factor values instead of item ratings. Finally, the impact of measurement errors for a single variable can be diminished by aggregating data measured on several similar variables.

Prior to performing a factor analysis, the data must fulfil some standards in terms of the quality and the quantity of the data. When calculating a factor analysis the resulting factor structure should represent the ‘real’ factor structure found in the population. In order to make such statements with sufficient certainty, a minimum sample size is needed. Guadagnoli & Velicer (1988, cited in Bortz, 2005) used a Monte Carlo simulation with subsequent regression analyses to determine this minimum sample size. Bortz (2005) transposed the resulting formula which can be used for this purpose and which is given in Appendix A 1.5.

Once the data have been collected but prior to conducting the factor analysis, an initial preliminary test of the item quality can be conducted by calculating the item difficulty. In the case of categorical items with several response options, the items must be recoded where necessary so that the lowest item value equals zero (Bortz & Döring, 2006, p. 219). According to Bortz & Döring (2006) item facility is calculated as:

$$p_i = \frac{\sum_{m=1}^{n} x_{im}}{k_i n} \tag{6}$$

where:

p_i = item facility (probability of item solution) for item i

x_{im} = item score of item i for single participant m

k_i = maximum item score for item i

n = number of participants in the sample.

Although the selection of items according to item facility depends also on the purpose of the test, a rule of thumb is that items with medium difficulty between 0.2 and 0.8 are preferred (Bortz & Döring, 2006, p. 219).

Factor analysis is based on the idea that the correlation between item value and factor value can be calculated. An item belongs to the factor with which it has the highest correlation coefficient. This correlation coefficient is termed ‘factor loading’ in the terminology of factor analysis. This seemingly easy method has a fundamental shortcoming: the factors and factor values are not known because they are the very result of the factor analysis itself. Therefore, these factor values must firstly be identified.

Statistically this is done by analysing the correlation matrix of the variables. The correlation coefficient between two variables x and y with the respective standard deviations s_x and s_y is calculated as

$$r = \frac{\operatorname{cov}(x, y)}{s_x s_y} \tag{7}$$

Concerning the rules used to interpret the value of correlation coefficients, values between 0.4 and 0.6 are regarded as medium, values between 0.6 and 0.8 as strong, and values above 0.8 as very strong (Brosius, 2008). In the case of *z*-standardised values (M = 0; SD = 1, see preceding chapters), which are used for factor analysis, the correlation equals the covariance. The covariance is calculated as:

$$\operatorname{cov}(x,y)=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{n}, \qquad (8)$$

where n is the number of observations, x_i and y_i the values and $\bar{x}$ and $\bar{y}$ are the mean values of the respective variables. Within this correlation matrix, the diagonal values which represent at this stage the correlation of an item with itself constitute a special case with important effects on further analysis. These diagonal slots must be filled with the communalities. When the analysis has been performed, these communalities represent the amount of explained variance of an item. However, at the beginning of the analysis, these commonalities have to be estimated. This estimate depends on the extraction method used. Therefore, the researcher has to decide which extraction method to use.

The extraction methods differ concerning the assumptions of how much variance can be explained by the factor solution. The most frequently used procedure is Principal (or Principle) Component Analysis (PCA), simply because it is the standard procedure in SPSS. However, strictly speaking this technique is not a factor analysis as it does not assume a causal relationship between factor and item (Bühner, 2006). 'Real' factor analytic techniques are Principle Axis Factor Analysis (PAF) and Maximum-Likelihood-Factor Analysis (ML).

PCA assumes that 100% of the variance of an item can be explained given that enough factors are used. Therefore, the communalities are set to one. PAF and ML assume that the factor solution can best account for this share of the variance of an item which can be explained by all other items. Thus, when a decision is made in favour of PAF or ML as the extraction method, the communalities are set to the squared multiple correlation coefficient (R^2) which is calculated by using regression analysis. According to Russel (2002, cited in Bühner, 2006) PAF is better suited to estimate the 'true' correlation matrix in the population. Where the assumptions of ML are met, ML is even preferable to PAF. However, as the ML assumption of multivariate normal distribution cannot be calculated in SPSS, but requires statistical packages such as AMOS (Bühner, 2006), ML is rarely used as a standard factor analytic method. In contrast, PCA requires neither normal distribution of the item values nor interval scaled values (although both are preferred) (Bühner, 2006).

The completed correlation matrix can then be analysed in order to further determine the suitability of the data for factor analysis. Three different criteria are usually used to quantify data suitability: the 'Measure of Sampling Adequacy' (MSA), the Kaiser-Meyer-Olkin criterion (KMO) and Bartlett's test of sphericity.

The KMO-value is a measure which is derived from the matrix of correlation coefficients and partial correlation coefficients of all items. The higher the correlation coefficients and the lower the partial correlation coefficients, the better the data is suited for factor analyses. This combination of correlation coefficient and partial correlation coefficient indicates that a third variable, which is to be identified later in a factor analysis, is responsible for the correlation between the two variables. The KMO-value sums up both coefficients across all possible item combinations (with the exception of self-correlations) and calculates its relationship. The closer the resulting KMO-value is to the maximum KMO-value of one, the better the data is suited for a factor analysis.

The second measure of data-suitability for factor analysis is Bartlett's test. This test is a χ^2 test which tests the null hypothesis that all items belong to an item population in which the correlation between variables is zero. Significant deviations from this null hypothesis can be interpreted as further indication of data-suitability for factor analysis (Backhaus et al., 2006; Brosius, 2008). It is important when interpreting the results of Bartlett's test to note that this test is sensitive to violations of the assumption of normal distribution (Backhaus et al., 2006; Bortz, 2005).

The MSA value can be interpreted as 'partial' KMO-value which is used for every single item instead of the entire correlation matrix (Brosius, 2008). It is therefore not identical to the KMO-value as suggested by Backhaus et al. (2006). Nevertheless, due to the close relationship between both criteria, the same rules apply with regard to their interpretation (Brosius, 2008).

The factor analysis can then be calculated by further analysing the correlation matrix. This correlation matrix is a square matrix which allows the use of the methods and laws of linear algebra or more specifically, matrix calculation. In this matrix, the correlation coefficients in each row or column are seen as elements of a vector in an n-dimensional space where n equals the number of variables used. The correlation matrix is decomposed into two matrices of Eigenvalues and Eigenvectors by Eigen-Decomposition (for a detailed description, see Borg & Staufenbiel, 2007; Bühner, 2006; Pett et al., 2003). The length of the Eigenvector is the Eigenvalue. The length of a vector (in German: 'Betrag') is calculated as the root of the summed squared vector values (Pythagoras-theorem). Applied to the correlation matrix this is:

$$z_{\text{length}} = \sqrt{\sum_{\text{row-wise}} z_i^2} \tag{9}$$

with z^2 being the column-wise sum of the squared correlation matrix including commonalities (see above). Each element of the vector is then standardised so that its length equals one, which is done according to:

$$z_{i\ \text{standard}} = \frac{z_i}{\sqrt{\sum\limits_{\text{row-wise}} z_i^2}} \tag{10}$$

The correlation matrix is subsequently multiplied by this standardised vector and the steps above are repeated until a certain stop criterion. This stop criterion is met when the difference of the summed squared differences between two consecutive vectors is less than .00001 (Bühner, 2006). In this case the solution 'converged' and the length of the last non-normalised vector (z_{lenght} see above) is the first Eigenvector in the principal component analysis. The factor loadings are calculated by multiplication of the vector values (equalling the column-wise sum of the correlation coefficients, see above) by the root of the Eigenvalue.

A residual matrix is built and the steps described above are repeated to find the next Eigenvalue until a pre-defined criterion. Depending on this criterion, the number of factors in the final factor solution might vary. In the literature, different criteria are given, amongst which are Eigenvalues > 1 or a decision according to the **scree-plot** of Eigenvalues. The decision of which criterion to use and how strictly to apply it depends also on the theory-based assumptions of the researcher. A discussion of the two criteria named above together with additional criteria is given in Bühner (2006).

Once the analysis has been finished, the final communalities of the items and the Eigenvalues of the factors can be calculated from the final matrix resulting from the correlation between item values and factor values (= the factor loadings). The commonality of an item represents the amount of explained variance by this item in the data and can therefore serve as an additional indicator of the suitability of an item in the factor solution. In the case of PCA and Varimax (orthogonal) rotation, the commonality h^2 of an item is calculated by summing the squared factor loadings of this item across all factors (the last column in Table 3 and Table 5). As the rotation solely serves to change the distribution of variance between the factors, commonality remains unchanged by the rotation. According to Bühner (2006, p. 192) the commonality can be seen as proxy for reliability in the case of single measurements. In this case, h^2 should be > .60, although such values are rarely met in reality (Bühner, 2006).

While the row-wise sum of the squared item loadings provides the explained variance of an item, the column-wise sum of the squared item loadings can be transformed in a measure of the explained variance of each factor and the overall factor solution. The column-wise sum of the squared item loadings results in the

Eigenvalues. In contrast to the communalities the Eigenvalues change with rotation which is why the Eigenvalues following rotation must be used. In the case of PCA, where the commonality of each item is set to one, the maximum variance equals the number of items. This is because PCA assumes that the entire variance can be explained by the factor solution (see above). As the variance of a standardised item equals one, the commonality of this item is one. Therefore, the percentage of explained variance of the factor equals the Eigenvalues divided by the number of items, multiplied by 100. Summing the explained variance for each factor gives the total explained variance of the factor solution.

Before the matrix is interpreted in terms of factor meaning, the quality of this entire solution must also be checked. One indicator is the internal consistency of a factor as a measure of reliability which can be estimated with Cronbach's Alpha. In short, high internal consistency means that the total test score can be approximated by a single item score. The higher Cronbach's Alpha, the higher the internal consistency, where a value of 0.8 is often given as the lower limit (Brosius, 2008). However, it should be taken into account that Cronbach's Alpha increases with the number of items (Cortina, 1993). Additional reasons supporting a careful interpretation of the total value of Cronbach's alpha are given in Shevlina, Miles, Davies, & Walke (2000). More important for this thesis is the comparison of alpha calculated with all items versus alpha calculated without an item of interest. If alpha increases without the item in comparison to alpha with the item, the use of this item in the test must be questioned. In the case of negative loadings on a factor, the items must be recoded before calculating alpha.

Another quality criterion for factor solutions is the corrected item-total correlation ('Trennschärfe' in German). These values denote the correlation between an item and the total score of the questionnaire without this item. As a rule of thumb, these values should not be below 0.3 (Field, 2009).

Finally, the factors can be interpreted. Items are assigned to a factor depending on the relative height of the factor loadings. The highest absolute value of the factor loading determines which factor the item belongs to. The factors can be interpreted according to the content of the items which are assigned to this factor. Within each factor the item importance for factor interpretation decreases with decreasing factor loading. The most important item for the naming and interpretation of a factor is the item with the highest factor loading on this factor (the so-called marker item).

In addition to this marker item, all other items with meaningful loadings must be considered when interpreting a factor. This is different to the calculation of factor scores for which all items are used (see below). The question in relation to the interpretation of these factors is how to define 'meaningful loading'. In fact, the difficulty in specifying such a value is reflected by the inconsistent specifications from different sources. These specifications range from 'above 0.3' (Comrad & Lee, 1992, cited in Pett et al., 2003), to ‘greater than 0.4’ (Stevens, 2002, cited in Field, 2009), ‘above 0.5’ (Backhaus et al., 2006), and ‘not below 0.6’ (Bortz & Döring, 2006). Comrad & Lee (1992, cited in Pett et al., 2003) based their interpretation on the shared variance which is derived by squaring the factor loadings. Their interpretation ranges from 20% shared variance which is regarded as fair (corresponding to a loading of .45) and 50% shared variance which is regarded as excellent (corresponding to a loading of .71). Alternatively, the interpretation of the correlation coefficients can be used as a rule of thumb because the loadings are in fact correlations themselves (see above).

After the factor structure has been determined, the aim is usually to perform further calculations. To do so, the factor scores are needed. A factor score is the value of a participant on each factor calculated from all its weighted item values. While in principle the weighting could be done by using the factor loadings it is usually done by using factor score coefficients. These are derived within a multiple linear regression approach. After standardization of the item values, these item values are multiplied by the factor score coefficients. The sum of these weighted item values is then the factor score. It differs from the calculation using the factor loadings in so far as different measurement scales are accounted for (Field, 2009). The resulting factor scores of both methods obviously differ from a mere averaging of item values as was done in the original RECL (see chapter 4.2.2). The difference is because the items in the factor score are weighted (see above) and because all items are included in each factor regardless of whether they ‘belong’ to this factor. Further advantages and disadvantages of the different methods of calculation of factor values are given in Bühner (2006), Pett, Lackey, and Sullivan (2003), and Backhaus et al. (2006).

4.2.5 Results

4.2.5.1 Replication of the RECL Structure

Prior to applying the RECL to validate the model it had to be assured that the factor structure found in the original RECL was replicated in the German version. Although the items could be averaged according to the original factor structure as was done by Steyvers, it was decided to perform a new factor analysis. This allowed replication of the findings with an even larger variety of rural roads while at the same time, new previously unknown but useful item combinations might occur.

The data was structured following Steyvers (1993) (see also Backhaus, 2006, pp. 326) which resulted in a 16 variable times 966 rating matrix (21 pictures times 46 participants). Following list-wise deletion of missing values there were 957 cases remaining for analysis. If the cases in the sample were participants, this sample size could almost be called ‘excellent’ (Bühner, 2006, p. 193). Applying the formula of Guadagnoli & Velicer (see preceding chapter and Appendix A 1.5) with an expected lowest factor loading of 0.54 found by Steyvers et al. (1994) and a factor stability of 0.9 would result in approximately 124 participants.

Table 3. Varimax normalised factor loadings of the RECL items after factor analysis.

Item	Factor I	Factor II	Commonality
Enjoyable	0.85	0.11	0.74
Dangerous	-0.79	-0.12	0.64
Threatening	-0.78	-0.07	0.62
Demanding	-0.77	-0.16	0.62
Gives a good view	0.76	0.31	0.68
Relaxing	0.73	0.25	0.60
Irritating	-0.70	-0.07	0.49
Spacious	0.69	0.31	0.58
Peaceful	0.66	0.08	0.45
Lowers concentration	0.16	0.86	0.76
Monotonous	0.11	0.85	0.73
Lowers alertness	0.19	0.80	0.68
Boring	0.03	0.79	0.62
Changeable	-0.06	-0.76	0.58
Inreases wakefulness	-0.34	-0.74	0.66
Increases attention	-0.38	-0.74	0.69
Eigenvalues before rotation	7.16	2.95	
Explained variance [%]	44.76	18.43	Sum: 63.19
Eigenvalues after rotation	5.42	4.70	
Explained variance [%]	33.85	29.34	Sum: 63.19

The assumption of normal distribution was violated, indicated by highly significant results of the Kolmogorov-Smirnov test (K-S test) for every item. However, as Bortz & Döring (2006, p. 218) state, the assumption of normal distribution is not relevant for large samples, which was the case here where the sample comprised 957 cases. In addition, the K-S test tends to become significant even with small deviations from normality where there is a high number of cases (Field, 2009). Furthermore, a visual inspection of the histograms revealed a reasonable distribution of the values.

In contrast to the K-S tests, the additional measures of data suitability were all satisfactory from the beginning. The KMO criterion had a value of .914 which is 'marvellous' (Kaiser cited in Backhaus et al., 2006, p. 276). Bartlett's test further indicated the suitability of the data ($\chi^2 = 10261.29$, df = 120, $p < .01$). The MSA-values were all above .8 with 11 of the 16 being above .9 and could thus all be termed 'meritorious' to 'marvellous' (Kaiser cited in Brosius, 2008, p. 780). The calculation of the item difficulty for the recoded item values of the entire sample resulted in values between .25 and .62 (*M* = .44, *SD* = .12), which is regarded as fulfilling the criteria (see preceding chapter).

The scree-plot of the Eigenvalues suggested using two factors (see Appendix A 2.4), whereas the height of the Eigenvalues would also have permitted a three factor solution. The Eigenvalue of the third factor was 0.94 and was thus only marginally below the Eigenvalue > 1 criterion which is often used to determine the number of factors (see preceding chapter). Using this third factor would have explained 5.88% of the variance before rotation and would have increased the total explained variance to 69.07%. With this Eigenvalue and in accordance with the three factor solution found by Steyvers (1993; 1998) a three factor solution seemed acceptable. This forced three factor solution is shown in Appendix A 2.5. Despite already deviating in some important aspects from the original RECL solution, it proved useful for a behaviourally relevant, self-explaining rural road categorisation (Weller, Schlag et al., 2008).

Due to the deviations and in order to minimise the factors, the two-factor solution was preferred here (Table 3). This is also regarded as more relevant for the central question which deals with identifying relevant parameters for explaining driving behaviour. Furthermore, the solution presented in Weller et al. (2008) showed some cross-loading items ('dangerous' and 'demanding'). Although these were of minor height and did not threaten the factor solution, a two factor solution might diminish these problematic cross-loadings.

This two factor solution explained 63 percent of the variance, with the two factors having an almost equal share after rotation. The values for commonality as additional indication of item quality showed satisfactory values with most being above .60. The items loading high on the first factor represented the entirety of items loading on Hedonic Value in the original RECL solution, whereas the second factor combined the remaining items which originally constitute Activational Value and Perceptual Variation. However, the polarity of these items was reversed in the solution depicted in Table 3. Regarding the naming of the factors, Hedonic Value could be retained for the first factor. In contrast, the second factor required renaming. In reference to the theory and by taking into account the polarity of the loadings, it is proposed here to name this factor 'Perceptual Monotony'.

4.2.5.2 Does This Factor Solution Allow the Prediction of Behaviour?

To assess the value of the two-factor solution shown in Table 3 to predict preferred speed for the different roads, the values for speed and the factor values were averaged road-wise. This resulted in a sample size of $N = 21$, representing the 21 road pictures. A linear regression was performed (method inclusion) with the two factor values on speed. This regression determines whether behaviour represented by rated speed can be predicted by perception represented by the factor values (Table 4).

Table 4. Results of the regression analysis of the averaged factor values on rated speed (laboratory study, method inclusion).

Variable	*B*	*SE B*	*β*	*t*	*p*
Constant	77.43			49.73	
Hedonic Value	17.64	3.14	0.55	5.63	.00
Perceptual Monotony	14.83	2.80	0.52	5.30	.00

Note. $R^2 = .87$, Corrected $R^2 = .86$, $F(2,18) = 60.05$, $< .001$.

Prior to performing the linear regression its assumptions were tested. The K-S test showed no significant deviations from a normal distribution for the variables used. The tolerance value was above 0.1 and the Durbin-Watson coefficient was only slightly above the limit of 2.5. Taken together, the assumptions were regarded as having fulfilled the criteria (see chapter 4.2.4.5).

The preliminary results support the model assumptions: both perceptual factors significantly contribute to predicting speed. This finding stresses the importance of a feed-forward, open-loop regulation of speed which is achieved through the perception of the entire road situation and exceeds the mere subjective representation of road geometry. Furthermore, the findings support holistic approaches of perception such as those of the Gestalt theorists and the direct approach to perception with the stress upon affordances. Nevertheless, the findings were termed preliminary because the factor structure of the RECL could not be replicated in comparison to the original RECL. However, such stability is needed for a practical applicability to rural road design. In order to identify potential commonalities despite different factor solutions, it is better to discuss the factor solution found with the simulator data here instead of in the chapter on the results of the simulator study.

4.2.5.3 Replication of the RECL Structure With the Simulator Data

RECL items were also collected in the simulator study which is the second study used for this thesis. In the simulator study, three straight road sections and nine curves were rated, from videos of the simulated course. Details of the simulator study including the sample and the course are all described in chapter 4.3.3.

Following preliminary tests with a small sample, some changes were made to the RECL compared to the version used in the laboratory and in comparison to the original RECL. The changes concerned the direction of some items and the number of anchor points used for the rating scales. With regard to the latter, a four-point scale was used instead of the six-point scale in the original RECL version; with regard to the former, 'increases alertness' was used instead of 'lowers alertness' and 'oppressive' was used instead of 'peaceful'. Furthermore, some changes were made concerning the order of presentation of the items (see Appendix A 2.2).

In accordance with the procedure described in Steyvers and Backhaus et al. (see chapter 4.2.2), the data for the simulator study was rearranged similar to the data in the laboratory study with all participants and picture-wise ratings in a single column for each item. With 12 rated elements (three straight road sections and nine curves) and 43 participants this resulted in $N = 516$ cases for each item. Due to a problem with the automatic data recording (see chapter 4.3.3.6), the ratings for half of the participants for two road elements were lost. Together with the deletion of single missing cases this resulted in a data matrix which contained 453 cases. Similar to the data in the laboratory study, such a sample size can be rated as between good and very good (Bühner, 2006, p. 193).

The different measures of data-suitability for factor analysis were as follows: the KMO-criterion had a value of .90 and can thus be regarded as 'marvellous' (see chapter 4.2.4.6). Except for the two items 'changeable' and 'monotonous' for which the MSA-values were .77 and .75, the MSA-values of all items were above .80 and can thus be termed 'meritorious' (see also chapter 4.2.4.6). With a significant result Bartlett's test also indicated the suitability of the data ($\chi^2 = 4285.57$, $df = 120$, $p < .01$). Again, the assumption of normal distribution was violated (highly significant results in the K-S test for all items) but this is again regarded as being a minor problem for large samples (see chapter 4.2.5.1). The analysis of item facility resulted in values between 0.2 and 0.8 for all items with most being around 0.5. Summarizing the results, the data were regarded as suitable for factor analysis.

A principal component analysis (PCA) rather than another extraction method was chosen comparable to Steyvers et al. (1994). The scree-plot (Appendix A 2.4) and the Eigenvalues suggested using a three-factor solution which is presented in Table 5.

This factor solution is more or less equal to the original RECL solution in regards to both the number of factors and the assignment of items to these factors. However, some important deviations were found. These concern the two items ‘demanding’ and ‘relaxing’ which are assigned to the first factor in the original RECL solution (termed ‘Hedonic Value’ by Steyvers). Before discussing questions concerning the potential implication of this finding, the issue is discussed from a methodological point of view.

Table 5. Varimax normalised factor loadings of the RECL items after factor analysis in the simulator study.

Item	Factor I	Factor II	Factor III	Commonality
Threatening	.82	-.13	.00	.70
Oppressive*	.81	.03	-.04	.66
Dangerous	.76	-.33	-.10	.70
Irritating	.76	-.26	-.02	.64
Gives a good view	-.64	.37	.12	.56
Enjoyable	-.59	.41	-.36	.64
Spacious	-.59	.27	.29	.51
Lowers concentration	-.11	.85	.14	.76
Lowers wakefulness*	-.13	.77	.25	.67
Increases attention	.33	-.76	-.20	.73
Increases alertness*	.42	-.72	-.13	.71
Demanding	.49	-.56	-.30	.65
Relaxing	-.53	.56	-.12	.60
Changeable	.14	-.13	-.87	.79
Monotonous	-.16	.11	.87	.79
Boring	.22	.40	.66	.65
Eigenvalues before rotation	6.91	2.46	1.36	
Explained variance [%]:	43.21	15.39	8.52	Sum: 67.12
Eigenvalues after rotation	4.52	3.80	2.42	
Explained variance [%]:	28.27	23.73	15.12	Sum: 67.12

Note. * Renamed items, see text.

Firstly, because of the similarity in the height of the factor loadings between the first factor and the second factor, it would be possible to forcefully assign both items to the first factor. In this situation the factor values would have to be calculated manually or the items could be averaged factor-wise as was done by Steyvers because such 'forced' assignment of items to factors is not possible in SPSS.

Further directions can be derived from the height of the factor loadings. Backhaus et al. (2006, p. 299) suggest only using factor loadings greater than 0.5 when interpreting a factor and further to use items which load greater than 0.5 onto different factors in the interpretation of each of these factors. Because the item loadings of both 'demanding' and 'relaxing' are either below 0.5 or just slightly above 0.5 one could propose to remove both items.

Such a procedure is also supported by Litfin, Teichmann & Clement (2000), and Bühner (2006) who in general propose deleting cross-loading items. These authors also state that the reason behind such cross-loadings could be inappropriate rotation methods. However, given the assumed orthogonal structure of the factors, the Varimax rotation as conducted in the original RECL is the appropriate method. Furthermore, changing the rotation method would not only change the loadings of the two items under consideration, but also the entire loading structure.

Another indication concerning the assignment of a certain item to a factor can be derived from calculating Cronbach's Alpha (see chapter 4.2.4.6). This was done for the two cross-loading items for the first and for the second factor. All items which had negative loadings onto the factors shown in Table 5 were reverse-scored factor-wise before calculating Cronbach's Alpha (see Field, 2009).

Table 6. Cronbach's Alpha and item characteristics for the two cross-loading items.

Item	Corrected item-total correlation	Alpha if item is deleted
	Factor I (Cronbach's Alpha: .899)	
Demanding	.638	.889
Relaxing	.655	.888
	Factor II (Cronbach's Alpha: .887)	
Demanding	.684	.870
Relaxing	.569	.888

Note. *N* = 7 items plus the two items 'demanding' and 'relaxing' for Factor I; *N* = 6 items for Factor II.

The results presented in Table 6 show that Cronbach's alpha calculated for the first factor did not increase if one of the two items 'demanding' or 'relaxing' was deleted. This means that both items could also be assigned to the first factor. In fact, the results for the second factor even suggest that the item ‘relaxing’ should be exclusively assigned to the first factor because Alpha increased after the item was deleted. This is also indicated by the corrected item-total correlation which is lower in this case than for the other items. However, it should also be stated that the changes in Cronbach’s Alpha are extremely small in all cases. Thus, from a statistical point of view both items could simply be deleted or assigned to either of the two factors.

As discussed in chapter 2.3.5 demand and the resulting workload are important concepts in existing motivational theories of driving behaviour. Simply deleting the item ‘demanding’ or arbitrarily assigning it to one or the other factor is therefore not an option, although it could be justified on statistical grounds. An alternative solution which would be in line with both the statistical analysis and the theoretical assumptions would be to interpret the item ‘demanding’ as kind of general ‘Demand’ or d-factor similar to the well known g-factor proposed by Spearman for intelligence (overview in Amelang et al., 2006). To identify such a d-factor, Principle Axis Factor Analysis (PAF; German: ‘Hauptachsenanalyse’) could be used (Borg & Staufenbiel, 2007). However, with the proposed d-factor being already represented in the data by the single item ‘demanding’ such a procedure does not make sense. Conclusions from the different factor analyses performed so far are made in the next chapter.

4.2.5.4 Conclusions Concerning the Factor Structure

From the preceding results it had to be concluded that the RECL factor structure did not show its assumed stability. Different factor solutions emerged from the different data sources and even within one data source. A decision in favour of a specific solution could not be made based on purely statistical reasons because all solutions introduced so far resulted in satisfactory values. Similar arguments apply to the content of the factors as defined by the items within the factors: no decision could be made in favour or against one of the proposed solutions because every solution ‘makes sense’ and can be supported by theoretical positions.

Although the solution found for the simulator data was at least similar to the original RECL solution, some deviations were found. These deviations could initially be regarded as minor; however, this is not the case. This is because they concern the item ‘demanding’ which is fundamentally important in driving theories. Furthermore, in comparison to the laboratory study, the stimuli used in the simulator study are of minor external validity. Moreover, they represent just a very small selection of rural road designs which in addition often only differed in a single experimental characteristic. From this perspective, the data collected in the laboratory study has a much higher relevance for future application to rural road design. However, it was also the laboratory data which showed most deviations in comparison to the original RECL.

Despite these seemingly discouraging results, one important aspect emerged from the different factor solutions which had not been focused on before in theories of driving behaviour. This is the importance of a factor concerning perceptual variation or monotony. This factor proved to be only slightly less important in predicting speed than the Hedonic Value factor which contains items such as 'dangerous' and 'demanding' (see chapter 4.2.5.2).

A final decision in favour or against one of the factor solutions presented in the previous chapters would require additional studies with a different sample to that used in the original RECL. However, in these further studies the stimuli should comprise a much larger variation of rural roads than just the two roads (albeit in different conditions) used for the development of the original RECL. Given these results it is not appropriate to use one or the other factor solution for further calculations. Therefore, a decision was made to instead select single marker items from the minimum two factor solution.

4.2.5.5 Selection of Marker Items

The rationale behind selecting marker items was to predict behaviour without the shortcomings of an unreliable factor solution as was presented in the preceding chapters. If items were not correlated, a selection of items based on their weights in a regression analysis on speed would be one appropriate statistical procedure. However, using regression analysis is not applicable to the question being discussed because on the one hand inter-correlation of items is a necessary prerequisite for factors to emerge in factor analysis (see chapter 4.2.4.6) but on the other hand multicollinearity of parameters in a regression analysis leads to arbitrary results (Brosius, 2008; Field, 2009).

Therefore, a selection based on the item loadings on the factors was preferred. This was at least possible for the factor, 'Perceptual Variation'. On this factor only three items consistently loaded throughout all factor solutions: 'changeable', 'monotonous', and 'boring'. 'Boring' was discarded because it consistently showed the lowest factor loadings. A preference for 'monotonous' over 'changeable' was made based on the factor solution presented in Table 3 where 'monotonous' showed higher factor loadings in comparison to 'changeable'.

For the remaining factor or factors, a selection of items merely based on item loadings on the factors was not appropriate. This is due to the instability of the factor solutions and to the differences in the loadings between different factor solutions. Therefore, a theory-based selection of items was used here based on existing motivational theories of driving behaviour (see chapter 2.3). As a result, the items ‘dangerous’ and ‘demanding’ were selected in a first step. Whether the number of items could be further reduced was analysed next. In fact there are some indications that the increase in information gained by using both items instead of using only one of them is minor. One indication is the close relationship between rated demand and rated feeling of risk as found by Fuller (see Figure 11). Whether this close relationship was also present for the data in this thesis was analysed by separately calculating the bivariate correlation between both items for the laboratory data and the simulator data. Because the data was not normally distributed (see preceding chapters), Kendall’s Tau was used in addition to Pearson’s correlation coefficient. The correlation coefficients can be regarded as between medium and strong for both studies, whereas squaring Pearson’s *r* resulted in only 35 to 41 percent of explained variance (see chapter 4.2.4.5):

- Laboratory study: $N = 46$; Pearson’s *r*: .64; Kendall’s τ: .59; both $p < .01$.
- Simulator study: $N = 45$; Pearson’s *r*: .59; Kendall’s τ: .55; both $p < .01$.

Given these results, there are differences between both items, however, these are minor in comparison to their commonalities and thus do not speak against a selection of only one of the two items.

Therefore, the two items were then analysed to determine which is better suited to predicting behaviour. For this step, a picture-wise (or road-wise) approach was chosen instead of using the combined data for all roads. This was done because the picture-wise pattern of the results might be more conclusive than only using the combined data. Again, bivariate correlations were calculated, this time between the item values of either ‘demanding’ or ‘dangerous’, and rated speed (Table 7). For this analysis, only the data for the laboratory study were selected because of the higher amount of road pictures and road characteristics.

The results indicate a slight preference for the item ‘dangerous’ which shows a higher number of significant correlations with rated speed (Table 7). Furthermore, ‘demanding’ is only once (road number 15) associated significantly closer to speed than is ‘dangerous’. This most likely means that ‘demanding’ cannot explain (Table 7 only shows correlations) behaviour on roads that cannot be explained by ‘dangerous’. Therefore, the item ‘dangerous’ was selected as the second marker item for the subsequent chapters. However, the rather weak coefficients indicate that neither dangerous nor demanding alone could sufficiently explain open-loop control of behaviour. This further indicates the appropriateness of selecting the second marker item ‘monotonous’ (see above).

Table 7. Road-wise correlation of rated speed and the items ‘demanding’ and ‘dangerous’.

	Item ‘demanding’		Item ‘dangerous’	
Road No.	*r*	*Tau*	*r*	*Tau*
1	-.02	.00	-.22	-.16
2	-.08	-.06	-.35 *	-.26 *
3	-.11	-.08	-.16	-.09
4	-.26	-.23	.10	.07
5	-.12	-.12	-.31 *	-.25
6	-.06	-.06	-.08	-.06
7	-.11	-.12	-.32 *	-.19
8	-.02	-.14	-.05	-.06
9	-.01	-.07	-.28	-.16
10	-.19	-.17	-.08	-.10
11	.23	.14	-.06	-.14
12	-.24	-.20	-.19	-.20
13	-.38 **	-.38 **	-.37 *	-.32 *
14	-.28	-.30 *	-.38 **	-.42 **
15	-.29	-.26 *	-.18	-.12
16	-.17	-.13	-.16	-.15
17	-.22	-.20	-.40 **	-.35 **
18	-.25	-.27 *	-.32 *	-.28 *
19	-.23	-.22	-.43 **	-.37 **
20	-.54 **	-.48 **	-.38 **	-.35 **
21	-.15	-.16	-.17	-.18

Note. $N = 46$ for each correlation.
* $p < .05$. ** $p < .01$.

4.2.5.6 Predicting Behaviour Through Perception Assessed With the Marker Items: Results for the Laboratory Study

As in chapter 4.2.5.2 a linear regression (method inclusion) was calculated on speed. In contrast to chapter 4.2.5.2, the averaged values of the two marker items 'dangerous' and 'monotonous' were used (Table 8) and not the averaged factor values. All values were averaged for each road in the sample resulting in N=21.

Table 8. Results of the regression analysis of the averaged item values for 'dangerous' and 'monotonous' on rated speed.

Variable	*B*	*SE B*	*β*	*t*	*p*
Constant	81.46			3.84	
'Dangerous'	-14.45	4.94	-.39	-2.92	.01
'Monotonous'	13.95	3.07	.61	4.55	.00

Note. $R^2 = .78$, Corrected $R^2 = .75$, $F(2,18) = 31.49$, $p < .001$.

Compared to the solution based on the factor values for the two-factor solution depicted in Table 4 and also in comparison to the three-factor solution depicted in Weller et al. (2008) (see Appendix A 2.5), the solution shown in Table 8 can be regarded as equally good. Again, it is concluded that differences in speed between different roads can indeed be predicted by the perceived behaviourally relevant characteristics of the road and the road environment.

4.2.5.7 The Role of Individual Factors

In the previous chapter the dominant role of the perception of behaviourally relevant road and environment characteristics was demonstrated. In this chapter, the role of individual factors on perception and subsequent expectations on appropriate behaviour is analysed. This was done to explore whether the role of individual factors in the model needs to be stressed further. In comparison to the road-wise analyses in the preceding chapter, the data had to be structured participant-wise for this task. This resulted in 46 cases with each case representing an individual participant. The dependent data in this matrix can either be averaged across all roads or used for several separate analyses for each road. In the former, potential differences between road categories might be levelled out. In the latter form of analysis, 21 separate results are generated which increase the probability of single results being influenced by chance (alpha-error accumulation, see chapter 4.3.3.15). Therefore, a combination of both approaches was regarded as appropriate: at first the data was analysed road-wise for individual differences and only afterwards were values averaged depending on the results of the first step.

As analysing individual factors was not the focus of the laboratory study, the analyses were carried out post-hoc with the limitations inherent to such analyses. In the case of the laboratory study, this mainly concerned the experimental selection of participants. Besides practical issues in relation to these shortcomings, the variable age was chosen for the analysis of group differences because of its well documented influence on driving (see chapter 2.2) and because of its relevance in relation to an increasing population of elder drivers (see also chapter 2.2). At first participants were assigned to three age groups according to the rules in Table 9.

Table 9. Classification rules for and descriptive statistics of the different age groups in the laboratory study [years].

Age group	Range	*n*	*min.*	*max.*	*M*	*SD*
I	up to 29	17	20	27	24.4	2.3
II	30 to 58	20	31	55	42.6	8.3
III	elder than 58	9	59	65	63.0	2.1

In order to obtain a reasonable number of participants in the third age group, a lower limit of 58 years was chosen. Because of the general decline of resources with age, it can be assumed that differences found between this age group and younger age groups would also be found if this limit was set at a higher age. Nevertheless, even the low age limit resulted in a low number of participants in the oldest age group and unequal subject numbers between the different age groups (see Table 9). The latter could be a problem for statistical analyses (see chapter 4.2.4.5) and should be taken into account when interpreting the results. A matching of participants was not performed as this would require that all relevant variables with regard to ratings of perception and behaviour were known. This was not the case here, owed to the new approach being used. Due to the limitations named above and because the analyses were solely a preliminary step towards further analyses (SEM, see below) differences were only analysed between the two extreme groups I and III.

To analyse these differences, road-wise (i.e. picture-wise) multivariate analyses of variance (MANOVAs) were performed with age group as the factor and expected appropriate speed and the two marker items ‘dangerous’ and ‘monotonous’ as dependent variables. The group-wise results for the K-S test indicated no significant deviations from the assumption of normality in the majority of cases (108 out of 126[8]). Box’s test resulted in non-significant results except for situation 11. Levene’s test was not significant except for ‘monotonous’ for road numbers 3 and 8, and speed for road number 4.

Table 10. Differences between the youngest and oldest driver groups with regard to speed and item values: significant results for the multivariate road-wise analyses of variance with the associated univariate statistics.

Effect of age group for road				Effect of age group for variable			
	$F(3, 22)$	p	Partial η^2		$F(1, 24)$	p	Partial η^2
No. 1	5.74	.01	.44	Speed	7.74	.01	.24
				Dangerous	3.04	.09	.11
				Monotonous	4.51	.04	.16
No. 5	4.44	.01	.38	Speed	9.90	.00	.29
				Dangerous	0.06	.81	.00
				Monotonous	2.92	.10	.11
No. 8	9.62	.00	.57	Speed	11.35	.00	.32
				Dangerous	0.37	.55	.02
				Monotonous	15.18	.00	.39
No. 15	4.53	.01	.38	Speed	7.30	.01	.23
				Dangerous	0.09	.77	.00
				Monotonous	3.39	.08	.12
No. 16	2.80	.06	.28	Speed	5.21	.03	.18
				Dangerous	2.19	.15	.08
				Monotonous	1.74	.20	.07

Overall, these results were regarded as having fulfilled the preconditions required for using MANOVAs for the question at hand. Significant and near significant results of the road-wise comparisons are shown in Table 10 in which the road-number is indicated in the first column. In general, only four out of the 21 roads showed

---

[8] 2 age groups × 21 roads × 3 variables (‘speed, ‘dangerous’, ‘monotonous’).

significant differences between the youngest and oldest driver groups with road number 16 approaching significance (see Table 10).

However, when differences were found, the effect size was quite large with the overall age-effect accounting for up to 57% of the variance. The differences in the multivariate tests can be attributed to differences in rated speed and differences in the ratings of the item 'monotonous'. Specifically younger participants reported a higher appropriate speed than older drivers and furthermore they rated these roads as being more monotonous (Figure 15).

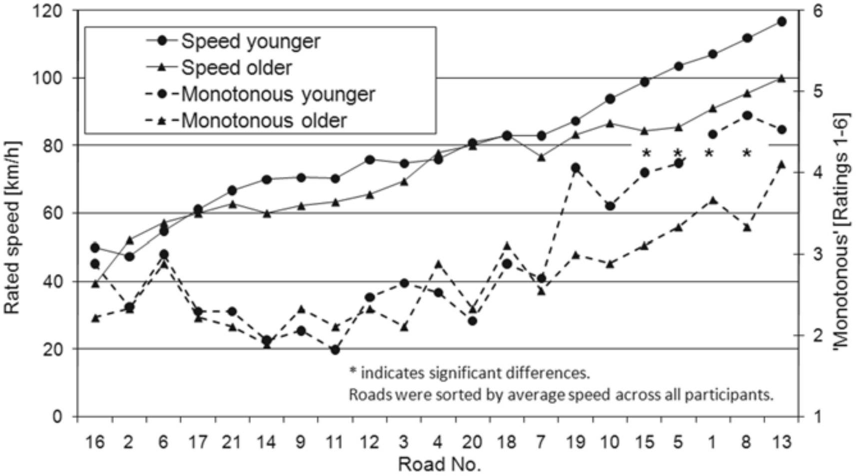


Figure 15. Road-wise differences between youngest and oldest driver groups for speed and ratings of 'monotonous'.

In addition to the significant differences in Table 10, the large differences for 'monotonous' between the two age groups for roads number 19 and 10 (see Figure 15) were tested separately with *t*-tests for independent samples. These differences were significant for road number 19 ($T = 2.13$, $df = 24$, $p = .04$) but not for road number 10 ($T = 1.41$, $df = 24$, $p = .17$). It is striking that the significant differences in the ratings of monotony as well as speed were all found for roads for which rated speed was very high for all participants. These were wide roads with very high sight distances, low curvature, good road markings and good surfaces. With the exception of road number 19 which is a tree-lined road, these roads were also all grouped in the same cluster in a preceding analysis of the data (see Weller et al., 2008).

The pattern depicted in Figure 15, indicates the possibility of interaction effects between road and age. To test this assumption statistically, the values of speed and the item ‘monotonous’ were averaged, once for high speed roads and once for lower speed roads. The breaking point below and above which seven roads were averaged was set at road number 19. Road number 19 was also assigned to the high speed roads because it was the first road in Figure 15 which showed significant differences in the ratings of ‘monotonous’ (see the t-test results above). Two repeated-measure analyses of variance were performed separately for speed and for the ratings of ‘monotonous’ with age group as the factor with the two levels age group I and age group III (Table 11).

Table 11. Results of two ANOVAs for ‘monotonous’ and speed in order to test differences between the youngest and oldest driver groups and between high and low speed roads (see text).

	$F(1, 24)$	p	Partial η^2
	Results for the averaged ratings of ‘monotonous’		
Road	92.62	.00	.79
Age group	3.85	.06	.14
Road × age group	10.83	.00	.31
	Results for the averaged ratings of speed		
Road	92.80	.00	.80
Age group	5.13	.03	.18
Road × age group	4.79	.04	.17

Again, preconditions were tested; however, they were not satisfactory for all tests. Specifically, the significant results for Box’s test indicated that homogeneity of covariances was not present. In addition, Levene’s test was significant for the item ‘monotonous’ for the slow roads (it was not significant in the other three cases). Owing to these restrictions a repeated-measures ANOVA was performed here, albeit for exploratory purposes only. The exploratory nature of this analysis is also due to the different sample sizes in the age groups (see Table 9 and chapter 4.2.4.5).

The most important result regarding the model and concerning the interpretation of the data is the significant interaction effect for both variables (see Table 11 and Figure 16): younger drivers interpreted faster roads as more monotonous and more conducive to faster speeds than older drivers. If this effect could be replicated with all statistical assumptions being met, it would indicate that younger drivers are particularly susceptible to a safety critical interpretation of high speed roads.

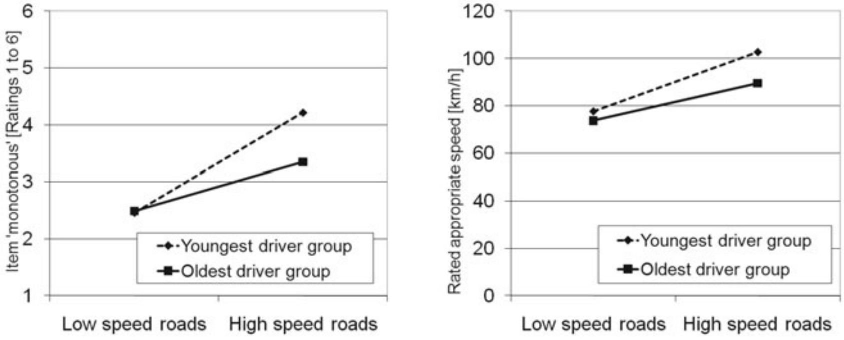


Figure 16. Averaged values for the ratings of 'monotonous' (left) and speed (right) separated by the youngest and oldest driver group and by low speed and high speed roads.

The results are important for the model in so far as they indicate that the influence of individual factors such as age depends on the characteristics of the road. In addition, the similarity in the pattern found for the ratings of 'monotonous' on one hand and behaviour on the other hand can be interpreted in favour of one of the model assumptions (with all due care): the perceived road characteristics determine the expectations regarding appropriate behaviour.

Although the results of the preceding analyses are subject to further statistical validation, an attempt was made to further analyse the role of individual factors. Of special interest for the model was where and how exactly individual factors could be further integrated into the driver and driving behaviour model for rural roads (Figure 14). In order to overcome some of the shortcomings of the preceding analyses, the age of all participants was included as covariate. Before the actual statistical analysis, the matrix of the correlation coefficients (Pearson's r) was calculated for all participants ($N = 46$) between age, the values for the two marker items 'dangerous' and 'monotonous', rated appropriate speed and the values for the item 'demanding' (Table 12). 'Demanding' was included to determine whether the decision for 'dangerous' as a marker item is confirmed by the data for the individuals. Because the effect of age group was restricted to the fast roads (see above), all items, including rated speed, were averaged for the fast roads only.

Once again, including 'demanding' did not result in any additional information which indirectly supported the decision to only select one marker item from the dangerous/demanding factor(s). The highly significant correlations for the high speed roads between age and monotony as well as between age and speed (Table 12) allow the subsequent alternative interpretations with regards to the influence of age on speed:

- Age could *directly and indirectly* influence speed (partial mediation model, see James et al., 2006).
- Age could *only indirectly* influence speed via perceived monotony (complete mediation model, see James et al., 2006) as presently depicted in the driver and driving behaviour model for rural roads (Figure 14).

Table 12. Matrix of the correlation coefficients ($N = 46$) for age and several items, all averaged for the fast roads only.

	Dangerous		Monotonous		Speed		Demanding	
	r	*p*	*r*	*p*	*r*	*p*	*r*	*p*
Age	.06	.69	-.35	.02	-.41	.01	.00	.99
Dangerous			-.07	.64	-.33	.02	.66	.00
Monotonous					.22	.13	-.05	.74
Speed							-.33	.03

A comparison of these alternatives can be statistically analysed with a number of different statistical methods (James et al., 2006). However, structural equation models (SEMs) are the method of choice for application to the driver and driving behaviour model. This is because SEMs allow the inclusion of the second marker item ‘dangerous’ in the statistical analysis. Thus, the complete model as proposed by the findings so far can be tested statistically. Again, because the analysis is restricted to the fast roads only, the results only apply to these roads.

The research question to be answered was whether age as an exogenous variable directly or indirectly influences speed. The indirect influence of age on speed is via the two marker items ‘monotonous’ and ‘dangerous’ which represent the other exogenous variables in the model. Speed in the context of SEM is the endogenous, that is, the dependent variable. In contrast to the usual SEM (see chapter 4.2.4.5), all variables named so far are manifest variables and thus strictly speaking do not constitute a structural equation model. However, the model structure outlined above could also be written in SEM-terms by artificially introducing latent variables. In this case, ‘behaviour’ could be introduced as a latent endogenous variable which unambiguously determines the single manifest variable ‘speed’. In this case, the coefficient between both variables could be formally set to one and the variance of the residual of speed could be set to zero. The same could be done with the exogenous variables in the model. However, this would artificially increase the complexity of the model without adding additional value in its structure. This is supported by Schumacker & Lomax (2004) who state that ‘single indicator variables of latent variables are not recommended’ (p. 212). Thus, the simple structure as outlined above was kept for the statistical testing.

The alternative method had to be used to calculate minimum sample size because the actual sample size was less than 100 (see chapter 4.2.4.5). For the partial mediation model depicted in Figure 18, five times the number of parameters to be estimated resulted in $N = 40$ which was achieved by the actual sample size of 46 participants. The assumption of normal distribution was tested with the K-S test which showed no significant deviations from the assumption of normality. The K-S test was preferred to the AMOS statistics for normality because it was the standard test in this thesis and because the latter are also only approximations (Arbuckle, 1997, p. 74). Maximum likelihood estimation could be used because the data was normally distributed. As stated in chapter 4.2.4.5, the same data had to be used for the testing of the different models. In combination with the small sample size, the subsequent analyses should be regarded as preliminary input for future model adjustments rather than as statistical testing in its strictest sense.

Table 13. Results for the linear regression of age, ‘monotonous’ and ‘dangerous’ on speed (all values averaged for fast roads only).

Variable	*B*	*SE B*	*β*	*t*	*p*
Constant	122.07	12.28		9.94	.00
Age	-0.30	0.11	-.39	-2.85	.01
Monotonous	1.80	2.14	.12	0.84	.41
Dangerous	-6.68	2.36	-.37	-2.84	.01

Note. $R^2 = .31$, Corrected $R^2 = .26$, $F(3,42) = 6.20$, $p = .001$.

A standard linear regression was performed before actually calculating the SEM models necessary to answer the research question. This was done both with SPSS and with AMOS to demonstrate the commonalities and differences between both programmes.

As expected, the standardised regression weights (i.e. the standardised path coefficients) between the exogenous variables and the endogenous variable ‘speed’ calculated by AMOS are the same in Figure 17 as in Table 13, in which the SPSS results are given. Because the number of parameters in the model (4 variances and 6 covariances) equals the number of parameters to be estimated (4 variances, 3 covariances, and 3 path coefficients) there are no degrees of freedom. In the terminology of SEM it is thus a saturated or just-identified model (Byrne, 2001, p. 118) and no significance tests concerning its likelihood are performed.

Regarding the covariance between the manifest exogenous variables, only the covariance between age and the item values of 'monotonous' was significant in the AMOS model ($p = .03$). The other covariance coefficients were quite low and were not significant (in Figure 17, the standardised covariance coefficients are shown, which equal the correlation coefficients).

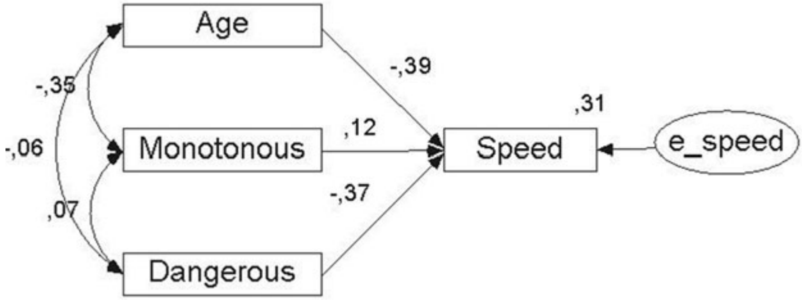


Figure 17. Path model of linear regression on speed with standardised coefficients.

The fact that only a single covariance was significant resulted in acceptable values for multicollinearity in the regression analysis: the smallest tolerance value (0.9) and the highest VIF value (1.1) were assigned to monotony but were both still in the acceptable range (tolerance much higher than 0.1 and VIF much lower than 10, see chapter 4.2.4.5). The test of autocollinearity which is summarised in the condition index resulted in a value of 19.2 at its highest and was thus an indication of moderate collinearity. The test of autocollinearity of the residuals provided by the Durbin-Watson statistics resulted in a value of 1.8 and was thus in the acceptable range between 1.5 and 2.5.

Based on these results (significant and non-significant covariances) and with respect to the research question, the partial mediation model was designed as shown in the path diagram depicted in Figure 18. Due to the fact that age influenced the ratings of 'monotonous' but not vice versa, the covariance between age and monotonous was transformed to a unidirectional regression path from age to 'monotonous'.

As a result, 'monotonous' itself became an endogenous variable and so had to be fitted with an error variable to account for the variance not explained by age. This model was compared to the complete mediation model depicted in Figure 19.

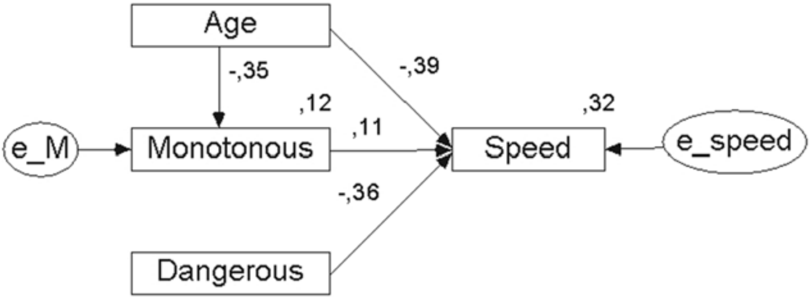


Figure 18. Partial mediation model for age and the ratings of monotonous on speed with standardised path coefficients and regression weights.

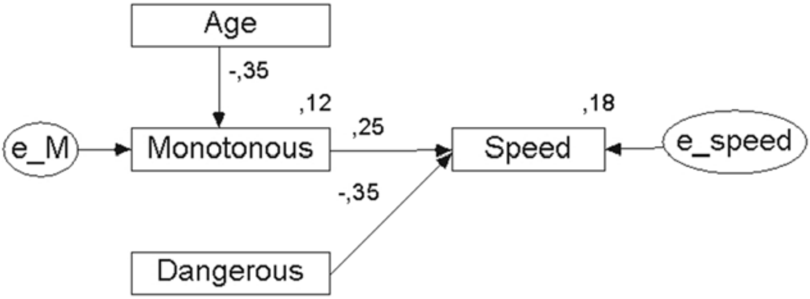


Figure 19. Complete mediation model for age and ratings of monotonous on speed with standardised path coefficients and regression weights.

The standardised regression coefficients of the independent variables on speed of the partial mediation model were the same as in the regression model (Figure 17). In addition, age explained 12% of the variance of 'monotonous' which was reflected by a significant path coefficient of -.35 (*C.R.*[9]= -2.48; *p* = .01). For the complete mediation model (Figure 19), the path coefficient of 'monotonous' on speed gained additional weight. However, it remained insignificant albeit only marginally (*C.R.* = 1.84; *p* = .07). By removing the direct influence of age on speed, the explained variance of speed decreased from 32% to 18%. The same results for 'demanding' instead of 'dangerous' would have resulted in 27% and 14% (not shown here), thus additionally supporting the decision to use 'dangerous' instead of 'demanding'.

However, the SEM approach was not calculated to maximise the explained variance of speed, but rather to test which of the two mediation models is more likely from a statistical point of view given the available data. The goodness-of-fit indices as introduced in chapter 4.2.4.5 are therefore summarised in Table 14, separated for the two alternative models.

Table 14. Different goodness-of-fit indices for the partial and complete mediation models.

Index	Partial mediation model	Complete mediation model
X^2	0.29	8.24
df	2	3
X^2 / df	0.15	2.75
p	0.86	0.04
GFI	1.00	0.92
AGFI	0.98	0.74
RMSEA	0.00	0.20
PCLOSE	0.88	0.06

All results exhibited a preference for the partial mediation model depicted in Figure 18 (for details concerning the interpretation of the values, see chapter 4.2.4.5). This was further supported by the standardised residual covariance matrix and the modification indices. For the partial mediation model, the highest standardised residual covariance was -.47 between the ratings of 'monotonous' and 'dangerous'. For the complete mediation model, the highest standardised residual covariance resulted in a value of -2.14 for the path between age and speed in addition to the ones for the

---

[9] The critical ratio C.R. is the estimated path coefficient divided by its standard error and should be > 1.96 in order to become significant (Arbuckle, 1997).

partial mediation model. This high value already indicated that the path between age and speed should be included in the model.

This interpretation was supported by the modification indices. For the two models, only the complete mediation model allowed the calculation of a modification index. This is because in this model the regression weight of age on speed is (implicitly) set to zero which is indicated by the missing path between both variables. The resulting modification index for this path was 6.40 with an expected parameter change in the χ^2-statistics of -.26. As the original value is assumed to be zero, this parameter equals the regression weight for this path if it were to be included in the model. It deviates from the value given in the partial mediation model (Figure 18), because the expected parameter change does not take into account other changes resulting from the inclusion of this path in the model.

4.2.6 Summary, Conclusions and Discussion of the Results

The laboratory study served several purposes. As a preliminary step, the structure of a German version of the original Road Environment Construct List (RECL) (Steyvers, 1993, 1998; Steyvers et al., 1994) was examined. This step was regarded as a necessary prerequisite for subsequent tasks concerning the validation of the driver and driving behaviour model for rural roads (Figure 14).

It was found that the original factor structure of the RECL could only partly be replicated. While the original factor 'Perceptual Variation' showed the expected item loadings and could be replicated with different data sources, this was not the case for the other two original factors 'Hedonic Value' and 'Activational Value'. Even the distinction between these two factors could not be reliably assessed in the laboratory study.

The reason for this discrepancy was seen as a result of the experimental stimuli used to collect the ratings. While the development of the original RECL was based on only two straight road sections, albeit in different traffic and light conditions, the data collected for this thesis was based on a large variety of two-lane rural roads. With respect to the subject of this thesis, the latter is favoured. Due to the results concerning the factor structure of the RECL, a decision was made to select only a few marker items from the entire list of potential RECL items. The selection of these items was based on both the statistical indicators provided by factor analysis and on considerations based on motivational theories of driving as introduced in the first part of this thesis. As a result, the two items 'dangerous' and 'monotonous' were selected.

After this preparatory analysis of the German RECL, the two marker items were used to test whether the perception of the road situation could indeed predict expected appropriate behaviour as assumed by the driver and driving behaviour model for rural roads. A linear regression analysis was used for the statistical test of this model assumption. The two marker items together with the constant explained 75% of the variance of rated speed which is interpreted in favour of the model assumptions. Furthermore, the result of the regression was of a similar quality to a regression performed previously with different factor solutions of the RECL and thus strongly supports the decision in favour of the two marker items. So far, these findings have two implications. Firstly, behaviour can be influenced by manipulating the 'look' of rural roads. Secondly, the selection of very few marker items provides a basis for how this manipulation should be conducted. For example, rural roads should not look monotonous. These findings were further supported in the subsequent simulator study where the effect of different designs was examined on measured speed (see following chapters).

Finally, the laboratory study was used to exemplarily test the model with respect to individual factors. These were so far only indirectly included in the model via 'Knowledge, experiences, and mental models' (see Figure 14). The research question to be answered was whether such individual factors should be additionally stressed in the model. Two alternatives were formulated: a partial and a complete mediation model (James et al., 2006). Both alternatives were formulated as path models and were tested statistically with the methods of structural equation models. The individual factor which was selected to test the alternative models was age, which is regarded as highly influential in determining driving behaviour (see chapter 2.2). Before actually testing the models, the data was tested for differences between the age groups. A significant interaction effect of age group with road type was found. Younger drivers rated fast roads as being more monotonous and gave higher speed ratings than drivers in the oldest age group. Thus, for the prototypical testing of the two alternative models, only the data for the fast roads was selected.

The results found were clearly in favour of the partial mediation model which assumes an additional direct path of individual factors on speed in addition to the indirect path via perception. However, at this stage it is not proposed to change the model by adding an additional path. This is mainly because the data which were used to test the alternative models were restricted to the subset of data for the fast roads. Even in this selected subset, the effect sizes were quite small. Furthermore, the division into age groups was conducted post-hoc, which resulted in the violation of several statistical assumptions and which consequently diminished the reliability of the results. Therefore, instead of adapting the model, further studies are proposed which focus on the influence of individual factors with respect to both perception and (expected appropriate) behaviour.

4.3 The Simulator Study: The Role of Cues and Affordances

4.3.1 Introduction: Rationale Behind the Simulator Study

The simulator study was conducted to examine the perception of both single cues and entire situations together with their effect on behaviour as proposed by the driver and driving behaviour model for rural roads (Figure 14). Using a simulator allows single elements to be unequivocally identified and unwanted external influences such as other traffic participants or different weather conditions to be ruled out. Thus, in contrast to the preceding laboratory study, the effect of specific elements on behaviour can be clearly identified. Furthermore, the simulator study is an extension of the laboratory study because behaviour is directly measured and not indirectly assessed by speed ratings.

4.3.2 Hypotheses and Additional Research Questions

According to the driver and driving behaviour model for rural roads (Figure 14) both affordances and single cues play a vital role in determining behaviour. The subsequent hypotheses were derived from the model and from the background introduced in the theoretical part of this thesis.

Hypothesis 1:

Speed is influenced by environmental characteristics.

Hypothesis 1a:

Adding objects to an otherwise barren environment will result in a reduction of speed.

This hypothesis was derived from the direct approach to perception (Gibson, 1986, see also chapter 2.4.4). According to this theory, velocity vectors play an important part in the perception of movement and speed. Perceiving speed in barren, monotonous environments is difficult because reference points are missing. Such reference points are necessary because they add velocity vectors to the field which – according to Gibson – are used by the driver to perceive speed (see chapter 2.4.5). If road geometry is kept constant, adding roadside objects should have a decreasing effect on speed, even if these objects do not serve as cues (see Hypothesis 1c). This hypothesis was tested on longer straight road sections in order to minimise the effect of road geometry.

Hypothesis 1b:

Speed on tree-lined roads is higher than speed on other roads.

In contrast to single trees and bushes placed at various distances away from the road, trees planted in a row alongside the road have a guidance effect on behaviour (Klebelsberg, 1982). This guidance effect is further accentuated by the gaze being directed towards the 'light at the end of the tunnel'. Both effects are caused because the single trees are perceptually combined into a single solid wall when the tree-lined road is driven through. Thus, no additional velocity vectors are provided and an increase in speed occurs (Gibson, 1986). Faster speeds on tree-lined roads (German: 'Allee') have already been measured in several studies (Zwielich, Reker, & Flach, 2001 for an overview). Thus, the results used to test this hypothesis can additionally serve to validate the simulator.

Hypothesis 1c:

Adding objects which serve as warning cues will result in a reduction of speed.

Almost all aspects of driving as explained in the theoretical part of this thesis can contribute to explaining the effect of cues:

- Cues help to guide attention to relevant locations or relevant aspects of the situation ahead (chapter 2.4).
- Cues are needed for open-loop control of behaviour (chapter 2.3.3) concerning both expected demand (chapter 2.3.5) and risk (chapter 2.3.4).
- Cues serve as valuable input for processes behind behavioural adaptation (chapter 2.3.2).

The direction of Hypothesis 1c is not only based on previous work (Driel, Davidse, & Maarseveen, 2004; Godley, Triggs, & Fildes, 2000; Manser & Hancock, 2007; Milleville-Pennel, Hoc, & Jolly, 2007) but to a larger extent on motivational models (see chapter 2.3). These explain driver behaviour as the effect of expected workload or risk. Warning signs increase perceived risk and demand and thus result in a corresponding decrease in speed. Hypothesis 1c thus directly leads to the next hypothesis.

Hypothesis 2:

Differences in speed are mirrored by differences in the perceived characteristics of the roads.

This hypothesis is directly derived from the driver and driving behaviour model for rural roads (Figure 14), which in turn is based on various existing theories of driving behaviour (see the first part of this thesis). Furthermore, this hypothesis was already exemplarily tested in the laboratory study (see preceding chapters). While cues in general might also directly influence behaviour without conscious processing (i.e. in a bottom-up way), the cues used in the simulator (e.g. warning signs, see chapter 4.3.3.6) will likely result in the hypothesised effects. With respect to the selected marker items (see chapter 4.2.5.5), the effect of objects merely enriching the environment versus objects possibly serving as cues might differ. However, at least one of the two items should show the hypothesised effects. A formal distinction is made between Hypothesis 2a and Hypothesis 2b in order to account for differences between straight road sections and curves.

Additional research question:

Do driver groups differ with regard to both perception of road characteristics and behaviour?

This question will be examined both for curves and straight road sections. The effect of driver groups is formulated as an additional research question rather than as a hypothesis. This is due to the fact that no specific selection of participants was conducted as group-specific differences were not part of the original model. Nevertheless, the results of the laboratory study and the literature (see chapter 2.2) suggest that such additional influences might be present. A test of the data for such differences, albeit post-hoc, is performed as an additional exploratory test of the model structure.

4.3.3 Methodology

4.3.3.1 The Experimental Paradigm Used

The experimental paradigm used in the simulator to test the hypotheses was a repeated-measures design with every participant driving every condition of the experimental factor. A single experimental factor was used which can be termed ‘road design’. This factor was subdivided for the statistical analyses into the two factors ‘curve’ (five conditions) and ‘long straight road sections’ (three conditions) (see chapter 4.3.3.6).

In addition to testing the hypotheses, differences between driver groups were analysed. Two-factorial analyses with both sub-factors concurrent (curves and straight road sections) were conducted only in relation to this additional question. The assignment of drivers to different groups was performed according to different demographic variables (see chapter 4.3.3.5).

The sequence of different conditions for the factor ‘road design’ was fixed (see chapter 4.3.3.3) and thus, strictly speaking, the design of the simulator study was quasi-experimental (Sarris & Reiß, 2005). However, because the simulated course was driven in two directions, the sequence of elements was balanced and thus allowed at least some experimental control of sequential effects (Bortz & Döring, 2006). Potential carry-over effects from one element to the next were analysed separately (chapter 4.3.3.12).

4.3.3.2 General Course of Events

The study was carried out in the driving simulator of the Fraunhofer Institute for Transportation and Infrastructure Systems IVI in Dresden[10]. IVI is situated directly adjacent to the campus of TU Dresden and can be easily reached by public or private transport. When the participants arrived at the simulator, the experiment leaders from TU Dresden and IVI introduced themselves and explained the general procedure. Following this introduction, the participants were asked to fill in a consent form and a questionnaire asking for biographical data, driving style and driving habits. In addition, a questionnaire developed by Richter, Debitz & Schulze (2002) was administered in which participants were asked for their present emotional state. This questionnaire was presented on a computer screen at the back of the simulator room (all questionnaires, see Appendix A 3.1).

Then, the control elements of the simulator car were introduced and adjusted to the participants. A test drive on a simulated rural road comparable to the experimental road was driven by the participants to allow them to get accustomed to driving in the simulator. Driving this test course took the participants approximately 10 minutes. When the participants finished the test course, the system was re-started and the participants drove the experimental course. The total driving time in the simulator was approximately 45 minutes which resulted from a combination of the experimental course and an additional road section that served other purposes. When the participants finished driving, they were asked to fill in the questionnaire of Richter et al. (2002) for a second time.

Afterwards, the second part of the study was conducted during which the participants had to rate videos of selected road elements of the experimental driving course. The ratings collected were a combination of the RECL items, including the two marker items (see the laboratory study), and additional items (see chapter 4.3.3.7).

---

[10] Further information is available at: www.ivi.fhg.de

Ratings were collected for all experimental curves and all experimental straight road sections (see chapter 4.3.3.6). In order to diminish effects of curve direction, all videos presented to the participants showed left curves, at least for the experimental sections reported here. To achieve this, the videos were taken in the outbound direction of the simulated driving course for some curves and in the inbound direction for others. Because both driving directions look the same when presented as single video clips, this is not regarded as having an influence on the ratings.

4.3.3.3 The Simulated Course

The unidirectional length of the experimental road section was approximately nine kilometres. After this experimental section, a section of approximately 15 kilometres was driven which served other purposes and is not reported here. At the end of this section there was a roundabout where participants turned around and drove the same course in the inbound direction, first the non-experimental section and then the experimental section. A bird's-eye view of the experimental road is shown in Appendix A 3.4; the curvature is depicted in Figure 20.

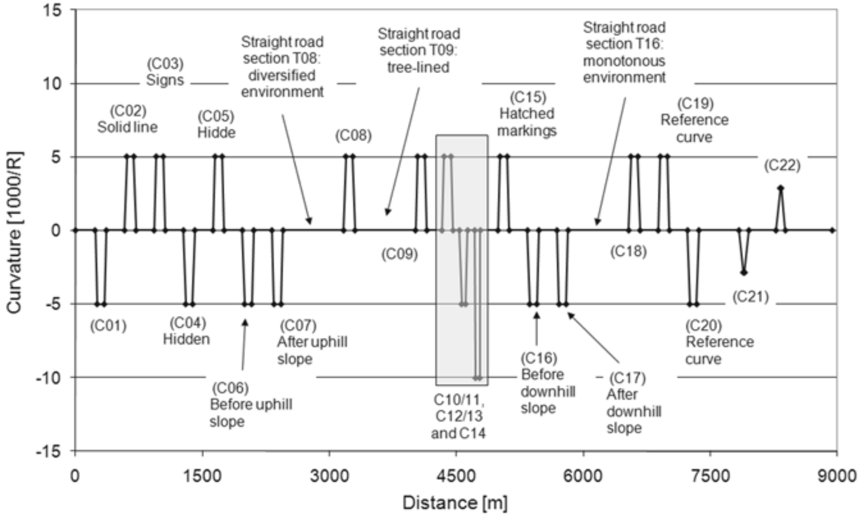


Figure 20. Curvature and distance of the experimental road section in the simulator. Curvature values below zero indicate right-turn curves when driven in the outbound direction. Slope direction is also given for the outbound direction.

The experimental road was programmed by IVI and contained 20 curves with short intermediate straight road sections (200 m) and three longer straight road sections, which were approximately 700 m long. The curves were programmed to have the same geometrical characteristics with a radius of 200 m and transition curves (clothoids) before and after the curve with a length of 25 m each. This resulted in a total curve length of approximately 130 m. Due to technical reasons and restricted resources, the two curves at the end of the course and the curves C10/11 to curve C14 (see Figure 20) are either of different geometrical characteristics or the intermittent tangents are of a different length compared to the usual intermittent tangents (see Appendix A 3.6 for details). Two slopes were programmed into the course which was uphill for the outbound direction between curves C06 and C07 and downhill between curves C16 and C17. Slope direction was reversed for the inbound direction.

After they had driven in the outbound direction, the participants drove the intermittent simulation of a real road (B6 near Dresden). The unidirectional length of this road was approximately 15 km. After a roundabout at its end, participants drove the entire course in the inbound direction. Further details regarding curve design and the design of the straight road sections are given in chapter 4.3.3.6.

4.3.3.4 The Sample

The sample consisted of 50 participants who were recruited in approximately equal numbers from IVI or TUD staff, and via an advertisement campaign in local newspapers. Because of simulator sickness (see chapter 4.3.3.10), this original sample had to be reduced to 43 people who were used for the analyses reported here. This remaining sample consisted of 27 males and 16 females with an average age of 38 for the male and 35 for the female participants. Average age for the entire sample was 37 with a range from 19 to 63 years (*SD* = 12.5). The average kilometres driven for the last three years were approximately 12,200 km/year (*SD* = 11,000) with ten participants also indicating that they had driven less than 5,000 km/year. All participants possessed a valid driving licence. Further information regarding the sample characteristics can be found in Voigt (2007). If participants were external (i.e. not TUD or IVI staff), they were paid to cover their expenses.

4.3.3.5 Assignment of Participants to Different Driver Groups

The additional research question (see chapter 4.3.2) suggested dividing participants into sub-samples. The division into groups was performed according to several variables. One important variable was age, due to its significant role in driving (see chapter 2.2). The limits of the age groups were chosen to assure comparable sample sizes in the different age groups. This, together with the fact that age was not the original focus of this thesis, resulted in a rather low bottom limit for the oldest age group of 48 years (see Table 15).

Table 15. Criteria for the division of the sample into age groups and descriptive statistics for age [years].

Age Group.	Range	*n*	*min.*	*max.*	*M*	*SD*
I	< 27	12	19	26	22.8	2.3
II	27 - 33	10	27	33	30.7	2.4
III	34 - 47	11	34	47	40.9	4.2
IV	48 <	10	48	63	54.5	4.6

Participants were further grouped by gender into 27 males and 16 females. In addition to these two variables, the participants were assigned to different groups according to their ratings in several items concerning driving style. These items were collected on a four-point rating scale ranging from 'does not apply at all' to 'applies in full'. Because the ratings were not equally distributed across all four categories, the two categories indicating agreement were assigned to a single category, as were the two categories indicating disagreement. This resulted in the subsequent distribution for the items used:

- item 'I like driving fast on straight rural roads' (21 drivers who did not like it versus 22 who liked it);
- item 'Compared to other drivers I tend to drive slower on rural roads' (28 drivers who disagreed versus 15 who agreed); and
- item 'How would you describe your driving style?' (28 'calm' versus 15 'sporty' drivers).

As mentioned above, differences between driver groups were not the focus of this thesis and were thus treated as an additional research question. The pre-selection of participants prior to the study was performed in order to control individual differences, not to test them. This pre-selection was necessarily limited to age and gender, and could not be performed according to driving style as there was no data available prior to the study. Further limitations concerning the analysis of differences between drivers are due to the different sample sizes in the sub-groups, which were a result of the post-hoc assignment to these subgroups. Due to the small sample size, testing interaction effects between different groups (for example between young males and older females) was not possible.

4.3.3.6 The Independent Variables

The assumptions formulated in the hypotheses were tested by comparing perception of and behaviour in road elements of different design. The road elements constitute the independent variables. For this thesis, a distinction was made between long straight road sections and curves. Three long straight road sections of approximately 700 m each were used with the following design of their environment:

- a diversified environment (T08);
- a tree-lined road (T09); and
- a monotonous environment (T16).

For the straight road sections, differences were tested between all three experimental sections. For the curves, differences with respect to cues were tested between the reference curve C19 without such cues and curves with cues (for the curve names, see Figure 20). The experimental curves of interest were

- curve C02, which had a solid, continuous middle marking instead of the usual intermittent line;
- curve C03, which was equipped with curve warning signs (signs no. 103 and 625 according to the German Traffic Regulations StVO);
- curve C05, with restricted sight distance (termed ‘hidden’ for short), which was achieved by placing bushes at the inside of the curve apex; and
- curve C15, where the lane width was perceptually narrowed by hatched middle markings.

4.3.3.7 The Dependent Variables: Initial Considerations

Two data sources were available for the subsequent analyses: subjective data from the ratings collected after the simulator drive, and objective data collected during the simulator drive.

Regarding the subjective data, the initial idea of using the RECL factor scores to assess the effect of different designs on perceived road characteristics had to be abandoned due to inconsistencies in the RECL factor structure (see the results for the laboratory study). Therefore, the two marker items chosen instead (‘dangerous’ and ‘monotonous’, see chapter 4.2.5.5) were used for the subsequent analyses. Additionally the item ‘demanding’ was used together with the item ‘the road element requires reduced speed’. The latter was used as a proxy variable for ‘Expectations concerning appropriate future behaviour for the situation ahead’ (see the model,

Figure 14). Other RECL items were additionally analysed, usually in order to rule out their having an additional influence which would otherwise have remained unnoticed.

Regarding the objective data, speed was selected as the relevant variable. This is because inappropriate speed is the most significant cause of accidents on rural roads which can be attributed to human error (Statistisches Bundesamt, 2007). Furthermore, the interrelationship between speed level and accident severity, and also the number of accidents, is well established (L. Aarts & Schagen, 2006; Elvik & Vaa, 2004). After this initial decision, further questions arose regarding:

- which is the relevant speed parameter (average speed, V85 or other parameters), and
- which is the relevant section for the calculation of this parameter?

The speed parameter used as the basis for statistical tests depends on the research question and the experimental design used. In road planning and road design, V85 is usually used as the relevant speed parameter. V85 is defined as the speed which is not exceeded by 85% of drivers in free-flowing traffic conditions on wet roads (FGSV RAS-L, 1995; Weise & Durth, 1997). Statistically it is calculated as the 85th percentile of the speed driven by different drivers at a given location. In Figure 22 an example of V85 values on a per metre basis is shown together with average speed of several drivers. While in principle the 85th percentile of speed could also be calculated for individual drivers (similar to an average speed within a certain road section), this is not what is meant by V85 as defined above. V85 as defined above results in one single value for each location or each road element. The use of such V85-values is shown in Figure 25.

V85 cannot be applied to test the hypotheses with the experimental setting used here because each experimental curve would only be characterised by a single value. Using V85 requires large sample sizes of road elements. The statistical test between two designs is then performed by comparing the V85-values of several curves of one design with the V85-values of several curves of an alternative design. Such approache is described in Dietze et al. (2005). In contrast to this report (Dietze et al., 2005), the study reported here trades a large sample size of road elements for high experimental control. Therefore, parameters on an individual basis are required such as average speed across a given road section, or the maximum or minimum speeds within this section.

Whether to use minimum or maximum speed depends on the road element and on the location of the measurement point. If only a single parameter is used for each element, it is useful to calculate the minimum speed for curves, assessed for the entire curve, and the maximum speed on straight road sections, assessed for the entire tangent. If more detailed analyses are required, a combination of minimum, maximum and average speed can be useful. This is the case when the approach behaviour to curves is to be described.

To statistically test differences between design variants, a section-wise calculation of these parameters can be used. A second possibility is to pre-define relevant behaviour or actions, such as the occurrence of a certain deceleration value or the first time the driver brakes. The location of these actions with relation to the curve beginning (or another characteristic point of the curve) can then be used as the parameter for the analysis.

A decision in favour of or against a certain parameter cannot be based only on theoretical considerations, but also requires a detailed pre-analysis of the data. This is especially the case in a simulator which has not yet been validated. Therefore, the final decision as to which parameter to use depends on the outcome of a detailed pre-analysis of the data which also takes into account aspects of simulator validity. This was done in the subsequent chapters, which resulted in the selection of a single appropriate parameter for the data at hand (see chapter 4.3.3.14).

Before proceeding to questions regarding speed, some remarks are needed concerning additional parameters. Besides speed, the lateral position of the vehicle on the simulated road was also recorded. However, the values of this variable turned out to be influenced by a non-systematic error between different locations. Although this shortcoming was recognised well before the experiments began, it could not be adjusted due to time constraints and a general shortage of resources. Instead, an attempt was made to recalibrate the recorded data to the data of a very experienced driver. This experienced driver drove the entire course at diminished speed and with the expressed intention of keeping a constant distance to the roadside. However, this provisional workaround was discarded after data inspection. The reason is that even this experienced driver had to rely on his perception of the driven scene and thus did not really provide an objective reference. Therefore, the variable 'position in lane' was not analysed further in this thesis.

With respect to the model evaluation, this shortcoming is regarded as being of minor importance. This is because the driving error 'leaving the lane', which is relevant for safety, is either due to excessive speed in curves (see Appendix A 1.3 for the physical relationship) or because of inattention or distraction. In the first case, assessing speed is enough to determine safety-critical behaviour. Inattention as the second reason for leaving the lane is usually due to sleep or to the driver being engaged in secondary tasks (Gordon, 2009). However, neither was the case in this study and subsequently no cases of drivers leaving the lane to the right were found during visual inspection of the experimental drives. Nevertheless, the variability of the in-lane position would have been an additional parameter of interest and should be included in future studies.

4.3.3.8 Ensuring Data Quality Prior to Data Analysis

Prior to defining the relevant parameters and statistically analysing the data, it was necessary to ensure that no systematic error diminished the validity and reliability of the data. This is of particular importance in simulator studies for which ecological validity is often questioned (see chapter 4.3.3.13). Besides this general issue, additional aspects had to be discussed for this thesis:

- the role of simulator sickness which resulted in the drop-out of participants (see chapter 4.3.3.4) and could have affected the performance of others who pretended not to have been affected;
- the issue of familiarity with driving in a simulator and the role of adaptation to it; and
- the influence of the order of the experimental road sections on behaviour.

Each of these issues was subsequently addressed in a separate chapter. In addition, a minor technical problem arose for the subjective ratings. Here, it turned out that the ratings of one curve (the curve with hidden sight) were only recorded for half of the participants. This issue was addressed first, followed by the issues relating to the objective data.

4.3.3.9 Issues Relating to Reduced Sample Size for Subjective Ratings

As was pointed out above, the sample size of the subjective ratings (not the objective data) was nearly halved to $n = 22$ for one single curve, namely the curve with restricted sight. This was caused by a programming error in the HTML-code used to present the questionnaire on the computer and to directly record the data in a data file. The error affected only half of the participants because the road elements were presented in reversed order to every second participant. The second curve which was affected by the programming error was not used in this thesis.

The analyses of the ratings in chapter 4.3.4.8 revealed that the reduced sample size in the ratings for this single curve might have affected the results. Therefore, a test was performed to establish whether the subsample of participants excluded from the data analyses due to data loss differed from the subsample retained for the analyses. If this was not the case, the pattern of results found in chapter 4.3.4.8 can be attributed to the reduced sample size alone. The factors chosen for this analysis were gender and age group as defined in chapter 4.3.3.5. This decision was supported by subsequent analyses of speed on straight road sections and in curves. For both elements significant and near significant differences were found between the participants when grouped according to the factors named above (chapters 4.3.4.2 and 4.3.4.7).

Prior to statistical testing the participant numbers were cross-tabulated for the groups defined above, separately for each factor. Chi-square tests were then used for the statistical analysis. Non-significant chi-square values indicate that differences are likely to be by chance only, meaning that both sub-samples probably belong to the same sample (see also Backhaus et al., 2006; Brosius, 2008; SPSS 16 Tutorial). The results for these tests were indeed not significant, neither for the factor gender ($\chi^2 = .26$; $p = .61$), nor for the factor age group ($\chi^2 = 3.00$; $p = .39$).

It can thus be concluded that the sub-sample of participants *with* data for the curve with restricted sight did not differ with respect to relevant factors from the sub-sample of participants *without* data for this single curve.

4.3.3.10 The Role of Simulator Sickness

Simulator sickness is a term which describes the negative effects of moving through virtual environments. Depending on the simulation, two kinds of simulator sickness can be distinguished: visually induced motion sickness (VIMS) and simulation or simulator sickness in its genuine form (Howarth & Hodder, 2008). Simulator sickness applies solely to moving-base mechanical simulation of movement. Because the simulator used for this study was a fixed-base simulator, VIMS is the correct term to be used. As with motion sickness, individuals differ in their susceptibility to simulator sickness or VIMS (Flanagan, May, & Dobie, 2005; Liu, Watson, & Miyazaki, 1999; Park et al., 2008). Simulator sickness, VIMS and 'common' motion sickness as experienced on ships or tilting trains are all supposed to have the same origins.

At present, conflicting theories exist and there is not yet a common understanding concerning their causes (Draper, Viirre, Furness, & Gawron, 2001; Flanagan, May, & Dobie, 2004; Schlender, 2008). Flanagan et al. (2004) have experimentally tested the reflexive eye movement theory, the sensory conflict theory, and the postural instability or ecological theory of motion sickness. Support was found in favour of the sensory conflict theory, but at the same time the validity of the other two theories could not be ruled out.

The sensory conflict theory of motion sickness is described by Reason (1978):

> All situations which provoke motion sickness are characterised by a condition of sensory rearrangement in which the motion signals transmitted by the eyes, the vestibular system and the nonvestibular proprioceptors are at variance one with another, and hence with what is expected on the basis of previous transactions with the spatial environment. (p. 820).

According to the sensory conflict theory, artificially simulating movement – either visually or mechanically – will aggravate motion sickness due to the inherent properties of a simulation. Amongst these properties are conflicting information through differences in the frequency of simulated motion versus self-motion (Duh, Parker, Philips, & Furness, 2004) as well as time delays. According to Draper et al. (2001), the latter might be of lesser importance. It is interesting to note that 'there is strong reason to suspect that increased realism may result in an increase in the incidence of simulator sickness' (Kennedy et al., 2003, p. 251).

The symptoms of VIMS and simulator sickness are the same as those for motion sickness. Kennedy et al. (1993) identified three distinct clusters of symptoms through factor analysis of a questionnaire: 'Oculomotor', 'Disorientation' and 'Nausea'. In a later study with the same items and non-military participants (some of whom had anxiety disorders; *N* total = 371), a two-factor solution emerged with 'Oculomotor' and 'Nausea' (Bouchard, Robillard, & Renaud, 2007).

These symptoms would make driving-simulators an inappropriate means of research if they occurred without a remedy. However, participants can be trained in the simulator resulting in habituation (Howarth & Hodder, 2008) and a decrease or even the disappearance of symptoms. In a study where participants played a computer game presented on a head-mounted display (HMD), Howard et al. (2008) found a floor effect for nausea after some training. The training consisted solely of the repetition of playing the computer game for 20 minutes in each training session. Whether the repetition took place on ten consecutive days or whether the ten training days were intermittent (not longer than for seven days) played hardly any role with respect to the effect of the training. Hu & Stern (1999, cited in Howarth & Hodder, 2008) report that the training effect lasts at least one month. Training in a motion-based driving simulator with emphasis on braking is described in Hoffmann & Buld (2006) with similar encouraging results. However, it is also estimated that despite training approximately 3% of participants will never habituate (Biocca, cited in Howarth & Hodder, 2008).

Such time-consuming training could not be conducted in this study due to a shortage of resources. The fact that seven participants (14% of the original sample) quit the experiment before having finished the course indicates that the simulator used does indeed induce VIMS. Because these participants were excluded from data analysis anyway, the central question is rather whether the participants who did not quit might nevertheless have been affected by VIMS.

In this thesis, this was not assessed by applying the Simulator Sickness Questionnaire developed by Kennedy et al. (1993), but more indirectly by using a questionnaire developed by Richter, Debitz & Schulze (2002). This questionnaire was developed for application in call-centres to assess the short-term consequences of work-related demand along four dimensions: positive engagement/well-being, fatigue, satiation/stress, and (emotional) monotony. The questionnaire consists of 12 items which have to be rated (‘At present I feel …’) on a six-point Likert scale with verbal anchors from ‘not at all’ to ‘very’.

Table 16. Differences in four factors of strain (Richter et al., 2002) before and after driving in the simulator: results of Wilcoxon tests for paired samples.

	Before		After			
Factor	*M*	*SD*	*M*	*SD*	*Z*	*p*
Positive engagement /well-being	4.05	0.91	3.42	0.88	-3.95	.00
Fatigue	2.20	0.99	2.91	1.09	-3.78	.00
Satiation/stress	1.51	0.61	1.50	0.64	-0.52	.61
Monotony	1.93	0.88	1.94	0.76	-0.01	1.00

Note. *N* = 43.

The values of the items which constitute a factor by Richter et al (2002) were averaged and tested for differences before and after driving. Because the values for satiation/stress were not normally distributed (significant K-S test), Wilcoxon tests for paired samples were used for the values of the 43 remaining participants in the sample on all four scales (Table 16). Despite the significant differences for positive engagement/well-being and fatigue (Table 16), the averaged values were still on the positive side of the scale. In combination with the non-significant differences for the other two scales, it can be concluded that the participants were still engaged and did not suffer from aversive symptoms. It was therefore concluded that symptoms of simulator sickness or VIMS did not affect the driving behaviour of the remaining participants in the sample.

4.3.3.11 Familiarisation With and Behavioural Adaptation to the Simulator

Ruling out simulator sickness or visually-induced motion sickness (VIMS) (see preceding chapter) could only be a first step towards reliable data. A second question concerned the familiarisation with driving in a simulator, more specifically the simulator used for this study. Such familiarisation is regarded as necessary because using the controls in a simulator differs from using the controls in real driving. This difference is because the input and output parameters of the mathematical functions behind simulated controls are only approximations of reality. Besides training to counteract simulator sickness as described above, participants are therefore usually familiarised with driving in the simulator.

In the literature involving driving-simulator studies, such a familiarisation process is often described in a very brief way, for example: 'participants were given a short practice run' (Mourant, Ahmad, Jaeger, & Lin, 2007, p. 146); '… short practice trips' (Dutta, Fisher, & Noyce, 2004, p. 219); '… practice session' (Broughton, Switzer, & Scott, 2007); 'participants were familiarised with the controls of the simulator' (Comte & Jamson, 2000, p. 142); were given 'a short training scenario' (Horrey, Wickens, & Consalus, 2006, p. 70); or 'naive participants received extra training' (Cnossen et al., 2000, p. 129).

Because the duration of such training is often shortened or extended depending on the participants, its duration is often only defined as 'until they [the participants] felt comfortable operating the simulator' (Charlton, 2007, p. 156). In other publications, the length of the practice run is given (for example, three kilometres, Rosey, Auberlet, Bertrand, & Plainchault, 2008), or the approximate time this process took. Here, the duration varies between studies, but is often given as 'a couple of minutes' (Verwey & Zaidel, 2000), which is specified as being usually between five minutes (Horberry, Anderson, & Regan, 2006; Horberry, Anderson, Regan et al., 2006; Yan, Abdel-Aty, & Radwan, 2008) and ten minutes (Farah, Yechiam, Bekhor, Toledo, & Polus, 2008; Gray & Regan, 2000). Some authors use longer time periods: Hoedemaker & Brookhuis (1998) used sets of 15 minutes, once with and once without ACC; Lenné, Triggs & Redman (1997) used 30 minutes.

Nevertheless, despite these practice runs, some authors report drop-out due to simulator sickness (Charlton, 2007; Yan et al., 2008) which indicates that these practice runs cannot replace anti-sickness training. However, for participants not susceptible to simulator sickness or VIMS, a duration of five to ten minutes should be enough. Such a claim is supported by McGehee et al. (2004) who found that steering movements in a simulator stabilised after approximately four minutes for naïve participants.

Besides the practice runs described above, insufficient familiarisation could, of course, also be controlled by balancing the order of the experimental conditions between participants or by repeating the conditions for the same participants (examples for both strategies are given in Charlton, 2007; Cnossen et al., 2000; Comte & Jamson, 2000; Dutta et al., 2004; Hoedemaeker & Brookhuis, 1998; Horrey et al., 2006; Lenné et al., 1997). Both strategies reduce the impact of potential insufficient familiarisation on data quality and results.

In the simulator study reported here, participants were familiarised with driving in the simulator prior to the experimental session. After the car controls were explained to the participants in stand-still mode, the participants were required to drive a test course of approximately ten kilometres. This test course consisted of similar two-lane rural roads with several curves as were present in the experimental condition. Because the simulated course was ‘hard-coded’ and because programming several versions of the course would have been too resource-consuming, the order of experimental elements in the simulated course could not be counterbalanced across participants. Therefore, the question of insufficient familiarisation gained additional weight.

Unlike the studies mentioned above, which could rely on the levelling effect of the experimental design, it was necessary to ensure before data analysis that experimental elements at the beginning were not influenced by insufficient familiarisation. Methodologically this could be done by comparing the data for the two directions in which the course was driven. Comparing the behaviour on the long straight road sections in particular was seen as being indicative of such effects. This is because curve direction does not play a role and because speed can be chosen freely according to the preferences of the participants.

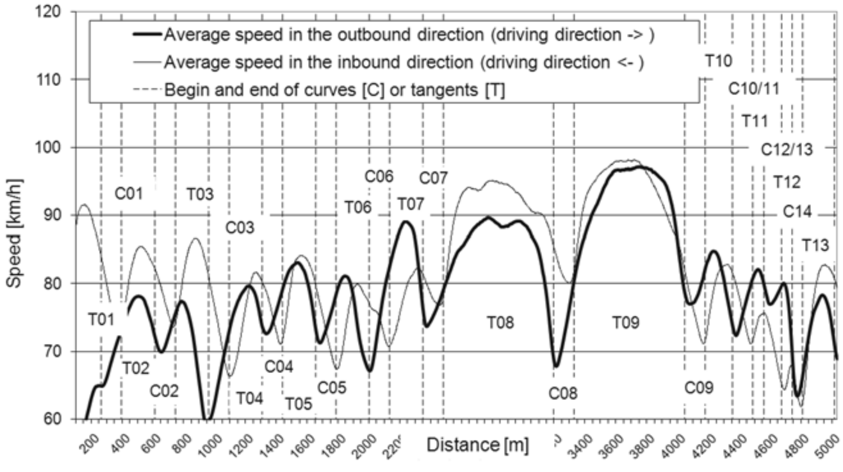


Figure 21. Speed [km/h] averaged across participants for elements C01 to C14, separated by driving direction.

In Figure 21 the data for both directions are illustrated by the averaged values per metre for the first (in the inbound direction: last) 5000 m of the experimental course. On the left side of Figure 21 large differences in average speed between the outbound direction and the inbound direction are clearly visible. However, it could not be determined whether this effect is attributable to insufficient familiarization or is an effect caused by stepwise acceleration processes towards a preferred speed. Regardless of its cause, the data prior to curve C04 had to be excluded from further data analyses. Another striking difference is visible at the long straight road section T08, throughout which speed was much lower in the outbound direction. The difference was tested statistically with *t*-tests for paired samples by using the parameter maximum speed throughout the entire straight road section. The same was done for the straight road sections T09 and T16 (Table 17).

Table 17. Differences in maximum speed between the outbound and inbound direction on the long tangents T08, T09, and T16: results of the t-tests for paired samples.

				Difference				
Section	Direction	*M*	*SD*	*M*	*SD*	*t*	*df*	*p*
T08	Outbound	93.12	13.08					
	Inbound	99.11	11.83	-5.99	9.71	-4.04	42	.00
T09	Outbound	101.06	13.23					
	Inbound	101.48	12.51	-0.42	8.51	-0.33	42	.75
T16	Outbound	101.17	12.65					
	Inbound	102.97	14.84	-1.80	9.01	-1.31	42	.20

The statistical results shown in Table 17 support the impression of the data shown in Figure 21: for the long straight road section T08, speed was lower in the outbound direction compared to the inbound direction. Especially for this section, it can be assumed that speed was determined by the speed on preceding road sections, and not a result of insufficient familiarisation. However, two aspects make such interpretation unlikely:

- First of all, the length of the long straight road sections was chosen to be independent of the preceding road sections (see next chapter).
- Secondly, if this effect was present, it should also have been found for section T09, before which speed was lower in the inbound direction.

Because no such effect was found for section T09 (see Table 17), it can be concluded that insufficient familiarisation indeed played a role, at least before section T09 in the outbound direction. Therefore, data before this section also had to be excluded from data analyses. Unfortunately, this decision affected the analysis of most of the experimental curves in the outbound direction. Analysing the remaining sections in the outbound direction would not have resulted in additional insight in the testing of the hypotheses. It was therefore decided to limit the testing of the hypotheses to the data for the inbound direction only. The fact that the data for the outbound direction and the inbound direction did not differ on the straight road sections T09 and T16, (i.e. after familiarisation) was seen as an indication that effects of road design were stable once familiarisation was achieved. Therefore, analysing data for the inbound direction only was considered to be sufficient to arrive at reliable and dependable conclusions.

At the end of this chapter, the question remains as to why participants needed such a long time to get accustomed to the simulation. This is even more astonishing because familiarisation with simulated car controls is usually achieved quite quickly and should have been finished at the end of the test course (see above). This is in accordance with my impression and with what participants expressed after the test course. It could be that the short break after the end of the test course caused by restarting the system for the experimental course (lasting around two minutes) might have played a role. The internal reference speed the participants developed during the test course might have been lost during this break because it was not yet stable enough. After the break, the participants might have needed to establish their internal reference speed anew. This can explain why each first encounter of a new situation (traffic signs, curves, long straight road sections) resulted in a decrease in speed which was not present the second time this situation was encountered. Finally, insufficient trust in the experimental situation in combination with the unfamiliar simulation might have played a role at the beginning.

Summing up this chapter, the findings again emphasise the need to only use participants for simulator studies who are familiar with driving that particular simulator and who are familiar with driving an experimental course in general. The resources needed to establish and maintain a database with a high number of trained participants would obviously be well spent. However, doing so requires human resources which cannot be made available as part of an ongoing research project with a limited time frame in which simulator studies constitute just one of several sub-tasks.

4.3.3.12 The Role of Element Order and Preceding Element on Behaviour

Simulator studies offer the possibility of examining experimental variations in controlled and safe conditions. The high degree of freedom in a simulator allows a simulated course and simulated events to be designed according to the researcher's will. On the other hand, these degrees of freedom can be a source of error themselves. Thus, the design of a simulated environment mirrors the researcher's present state of knowledge, concerning both issues specific to the topic being researched and experimental methodology in general. However, in reality, restraints due to a lack of resources can lead to fundamental differences compared to the original plan.

For the study at hand, resources were limited because the simulator study was not part of the project work-plan and was conducted supplementary to the original plan. As a consequence, some changes to the original plan had to be made and could not be avoided. Whether – and if so, how – these changes influenced the data had to be tested prior to data analysis, as with the influence of simulator sickness (chapter 4.3.3.10) and insufficient familiarisation (preceding chapter).

Originally, all curves were intended to have the same radius of 250 m. After the changes, the curves C14 (R = 100 m), C21 and C22 (both R = 300 m) deviated from the rest of the curves. As these curves were not part of the experimental variation, this is of minor importance for the study at hand. Similarly, the implemented change in radius from R = 250 m to R = 200 m for all other curves does not pose a major threat to the experimental quality because it is consistent throughout all curves. In contrast to this, the change in tangent length between curves might indeed be an issue. Originally, tangents between experimental curves were all planned to be 500 m in length, but were later reduced to approximately 200 m. Similarly, the long straight road sections were planned to be 1000 m but were reduced to approximately 700 m (see Appendix A 3.6 for the exact lengths).

The question of interest was whether these changes in tangent length had an unwanted influence on the data of the subsequent experimental curve. Because tangent length was shortened, the question can be formulated more precisely as: is there a carry-over effect of speed in the preceding curve to the speed before and in the experimental curve? To answer this question, speed prediction models are needed which take into account the effect of tangent length and curve characteristics. In fact, such approaches were developed in road engineering. Provided that behaviour in a simulator is comparable to real behaviour on the road, these approaches can also be used for the data collected in the simulator. Because the simulator had not yet been evaluated in terms of its external validity at the time of the experiments, comparing data in the simulator to 'calculated' behaviour from the engineering models can also be used as input for a preliminary 'proxy' validation. To do this, the engineering approaches are introduced first and then used later in chapter 4.3.3.13.

Before explaining a selected engineering approach in detail, Figure 22 is used to demonstrate the issue at hand. The section depicted in this figure is the non-experimental section including curve C14 which only had a radius of 100 m (see above). The different radii and the different tangent lengths in this section made it especially suitable for demonstrating the influence of the preceding elements.

An obvious example of the influence of the preceding road element can be seen in Figure 22 for the values of tangent T12 which had a length of only 80 m (see Appendix A 3.6). Depending on whether curve C14 was driven before or after tangent

T12, maximum speed on this tangent differed significantly (Table 18). Of course, this effect could also be attributed to insufficient familiarisation. However, it was established in the preceding chapter that such an influence did not play a role after section T09 and thus also not on tangent T12.

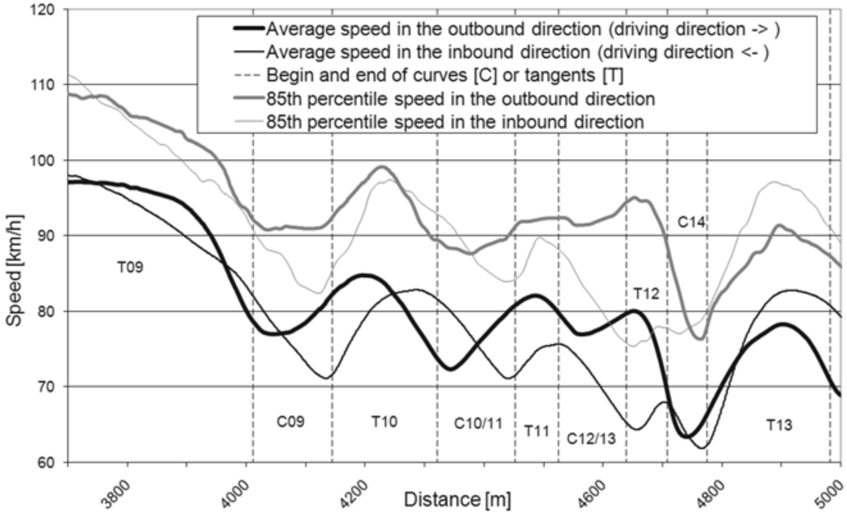


Figure 22. Average and 85th percentile of speed for curves C09 to C14, separated for the outbound and inbound direction.

Table 18. Differences in maximum speed on tangent T12 depending on whether curve C14 was driven after T12 (= outbound direction) or before T12 (= inbound direction): results of the *t*-test for paired samples.

				Difference				
Section	Direction	*M*	*SD*	*M*	*SD*	*t*	*df*	*p*
T12	Outbound	80.80	11.99					
	Inbound	69.68	9.08	-11.12	7.51	-9.71	42	.00

Although none of the elements shown in Figure 22 were later used as experimental elements, the example showed that determining an appropriate tangent length is essential in planning a simulated course. Before the situation for the experimental sections is analysed, an existing approach will be introduced which allows appropriate tangent lengths to be determined. As this approach has not been applied in simulator studies previously, its applicability in the context of the ecological validity of the simulation will be demonstrated in an extra chapter (chapter 4.3.3.13).

The first step in defining an appropriate tangent length is to determine speed for the geometrical elements to be used. For existing roads, speed can be predicted based on the geometric curve characteristics either by determining the maximum possible speed, or by measuring real speed in the field and subsequently generalising this data with approximation functions for other, non-measured curves.

In the first situation, maximum speed defines an equilibrium between resistant, adhering, and driving forces. In the 'Green Book' (AASHTO, 2001), these processes are described:

> When a vehicle moves in a circular path, it undergoes a centripetal acceleration that acts toward the centre of curvature. This acceleration is sustained by a component of the vehicle's weight related to the roadway superelevation, by the side friction developed between the vehicle's tires and the pavement surface, or by a combination of the two. As a matter of conceptual convenience, centripetal acceleration is sometimes equalled to centrifugal force. (p. 131)

How these laws of physics are applied in order to calculate a maximum possible speed for a curve with a certain radius is described in Appendix A 1.3. However, these formulas were developed to calculate the superelevation of a new road based on the known design speed. So strictly speaking, resolving the formulas for speed results in the design speed and not the maximum possible speed. For a given side friction, a given superelevation, and the radius R = 200 m used for the experimental curves, these formulas result in speeds between approximately 55 km/h for curves without superelevation and approximately 70 km/h for curves with 8% superelevation (see Appendix A 1.3.).

The second approach, which uses measured speeds in the field to approximate the 85th percentile of speed, allows a closer estimation of real speed. For a curve with a given radius, V85 can be estimated by using RAS-L (FGSV, 1995, Abb. 34, p. 39). For a radius of 200 m and road widths below 6.5 m, the speed given in RAS-L is approximately 90 km/h. V85 on tangents is equalled to the design speed plus 10 or 20 km/h (FGSV, 1995). Although determining the influence of the preceding element on the speed on tangents is not usually needed for road design, it is needed for safety analysis. Here, measured speed is combined with driving dynamic models. For example, Sossoumihen (2001) and Steyer (2004) used a distance of 400 m before the curve as the relevant distance to calculate an averaged weighted curvature, which can subsequently be used as input for the formulas described above.

Another, more flexible approach was developed by Lamm et al. (2007). This approach is regarded as superior to other existing approaches because it also allows a safety assessment of rural roads. It will be explained and its application will be shown for the simulated environment. This rather detailed description is provided because:

- the implemented simulator course differed from the planned simulator course, and thus required a precise estimation of the influence of these changes, and,
- as far as known, this process has never before been described or applied as part of designing a simulated driving course.

Similar to the approaches described above, Lamm et al. (2007) used measured speed as input for regression analyses which resulted in the following formulas. One is for Germany derived from German data and one is for 'worldwide' application derived from data collected in various countries.

The formula for calculating V85 in Germany is given in Lamm et al. (2007), as:

$$V85 = 10^6 / (8270 + 8.01 \times CCR_S) \tag{11}$$

and for 'worldwide' application it is given in Lamm et al. (2007), as:

$$V85 = 105.31 + 2 \times 10^{-5} \times CCR_S^2 - 0.071 \times CCR_S \tag{12}$$

For curves without transition curves, CCRs is determined as:

$$CCR_S = \frac{\alpha}{L_C} = \frac{63662}{R} \tag{13}$$

and with transition curves it is determined as:

$$CCR_S = \frac{63662 \times \left(\frac{L_C}{R} + \frac{L_{A1}}{2R} + \frac{L_{A2}}{2R} \right)}{L_C + L_{A1} + L_{A2}} \tag{14}$$

where:

CCRs = curvature change rate of the single curve [gon/km]

R = radius [m]

L_C = length of the circular curve [m]

L_A = length of the transition curve (clothoid) [m]

α = central angle and deflection angle [gon].

Further details regarding the geometric background of calculating CCR and CCR_S are given in Appendix A 1.4.

The formulas above can also be used to determine V85 for tangents. In this case CCR_S is equalled to zero. However, this pertains exclusively to long, so-called ‘independent’ tangents. A tangent is independent if its length allows drivers to accelerate from V85 in the preceding curve to V85 for tangents, and to decelerate from this V85 to V85 in the subsequent curve. Mathematically this length is determined as:

$$TL_{V85\max} = \frac{(V85_{\max})^2 - (V85_1)^2}{2 \times 3{,}6^2 \times a} \qquad (15)$$

where:

TL_{V85max} = tangent length needed to accelerate from V85 at curve one to the maximum speed on tangents V85max [km/h]

$V85_1$;V85max = 85th percentile speed in curve one and for long tangents [km/h]

a = acceleration assumed: 0.85 [m/s^2].

These processes are illustrated schematically in Figure 23, which allows a more general discussion than the actual values depicted in Figure 22. The starting point in Figure 23 is a given fictitious speed of 40 km/h for a curve located at distance zero. After the curve, the driver accelerates until the next curve requires the driver to decelerate or until a certain maximum speed is reached.

In Figure 23, this maximum speed was set to 100 km/h, which represents the maximum permitted speed for rural roads in Germany. According to Formula (15), and with the assumed acceleration/deceleration of 0.85 m/s^2, this point is reached at a distance of 385 m after curve one. Depending on the perceived appropriate speed of the next curve, drivers will decelerate from $V85_{max}$ to this speed at the distance required to do so.

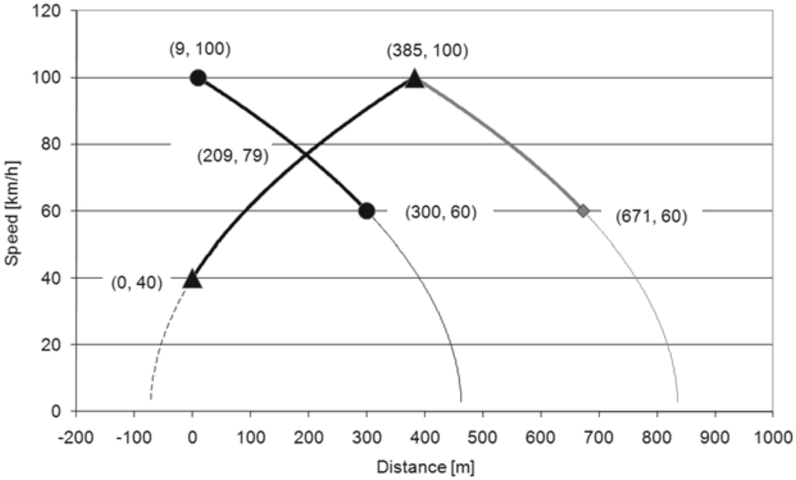


Figure 23. Schematic representation of velocity on two tangents of different length between two curves (further explanations in the text).

Again, to determine this distance, a value for deceleration must be assumed. The distance itself can again be calculated by using Formula (15). According to Lamm et al. (2007), Formula (15) and – depending on whether values are to be calculated for Germany or worldwide application – Formula (11) or Formula (12) can be combined to:

$$TL_{\max} = \frac{2 \times (V85_{T\max})^2 - (V85_1)^2 - (V85_2)^2}{2 \times 3{,}6^2 \times a} \tag{16}$$

where:

TL_{max} = necessary acceleration/deceleration length to reach $V85_{Tmax}$ between curves 1 and 2 [m]

$V85_{Tmax}$ = maximum operating speed on tangents for CCR_S = 0 gon/km [km/h]

$V85_1$; $V85_2$ = 85th percentile speed in curves 1 and 2 [km/h]

a = average acceleration / deceleration [m/s^2].

For all tangents longer than this TL_{max}, the speed $V85_{max}$ is assumed. In the example used in Figure 23 and with an assumed appropriate speed of 60 km/h for the second curve, the formula results in a distance of 671 m. If the tangent between the two curves is shorter, the driver will not reach the maximum speed (assuming that the same acceleration/deceleration values are used). The speed the driver will reach on shorter tangents can be calculated according to Lamm et al. (2007) as:

$$V85_T = \sqrt{2 \times 3{,}6^2 \times a \times (TL - TL_{\min}) + (V85_{1;2})^2} \tag{17}$$

where:

$V85_T$ = 85th percentile of speed on a tangent [km/h]

TL = actual tangent length [m]

TL_{min} = minimum tangent length needed to accelerate from one element to the next element [m]

TL_{min} is calculated according to:

$$TL_{\min} = \frac{\left|(V85_1)^2 - (V85_2)^2\right|}{2 \times 3{,}6^2 \times a} \tag{18}$$

where:

$V85_{1;2}$ = 85th percentile speed in curve one or two; for $V85_T$ always the maximum is used [km/h]

a = acceleration assumed: 0.85 [m/s^2].

For the example illustrated in Figure 23 with the second curve located at a fixed distance of 300 m after curve one, this formula results in a maximum speed of 79 km/h which is reached at a distance of 209 m after curve one.

After this introduction with fictitious data for the formulas of Lamm et al. (2007), the formulas were used to determine whether the tangents between the experimental curves in the simulated course were long enough to be regarded as independent. This is the case if the tangents are at least as long as TL_{max}. Only this case ensures that speed before the experimental curve is not influenced by speed in the preceding curve, and only in this case can speed data before the experimental curve be used to assess the anticipated behavioural adaptation to the experimental curve. Because of the central role of acceleration/deceleration for the outcome of the formulas, it was necessary to ensure that the acceleration/deceleration found in the simulation was comparable to the assumed value of 0.85 m/s², which is used both for acceleration and the absolute value of deceleration by Lamm et al. (2007). Acceleration is calculated according to:

$$a = \frac{{v_2}^2 - {v_1}^2}{2s} \tag{19}$$

where:

a = acceleration [m/s²]

v_1 ; v_2 = speed at location one, location two respectively [m/s]

s = distance between locations one and two [m].

Despite this seemingly clear-cut definition, a decision must be made regarding which data sets to use for locations one and two. This decision depends on thematic considerations. In principle, the location assigned to one data record and the location assigned to the next data record in the database could be used. This method is useful for ADAS applications which require constantly updated values. For this thesis, which deals with the influence of road elements on speed behaviour, the relevant locations have to be related to these elements. To do so, acceleration and deceleration were calculated between the locations for a maximum and a minimum speed.

Maximum speed needed to calculate deceleration was defined within a section from the start of the curve to 200 m prior. Maximum speed needed to calculate acceleration after the curve was defined from the end of the curve to 200 m after. The 200 m for the calculation of maximum and minimum were chosen because of the short tangent lengths in the simulator which were approximately 200 m. Minimum speed was defined for the road section covering 50 m before the beginning of the curve to the end of the curve. Those 50 m were added because in some curves some participants reached minimum speed before the actual beginning of the curve. The distribution of the values for curve C20 in the inbound direction is exemplarily shown in Figure 24.

Averaged across all participants and across all curves for the inbound direction, this resulted in a value of 0.48 m/s² for acceleration and 0.76 m/s² for absolute deceleration. The distribution of the values for the participants is shown in Appendix A 3.8. Although acceleration and deceleration differed significantly (see Appendix A 3.8) and although values were slightly lower than the 0.85 m/s² assumed by Lamm et al. (2007), the differences are regarded as minor, given that deceleration values of 5 m/s² can be reached when braking hard to a standstill. Thus, for the subsequent calculations, 0.85 m/s² was used as the input value for acceleration/deceleration.

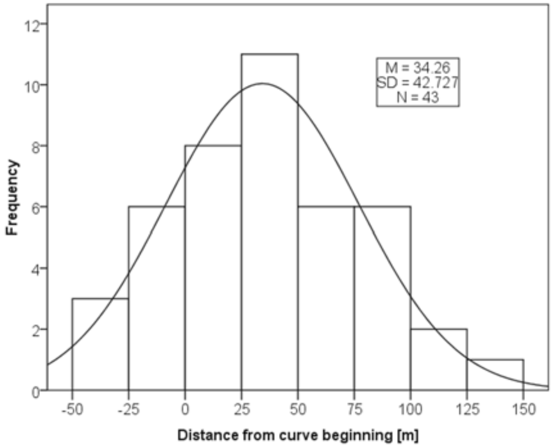


Figure 24. Exemplary distribution of distance from curve beginning where minimum speed was reached. Shown here: curve C20 in the inbound direction; curve beginning equals zero, a negative value indicates a location before curve beginning.

This value was used in combination with the CCR_S value of the experimental curves to calculate the minimum tangent length required between two curves in order to be able to assume – according to Lamm et al. (2007) – that maximum speed on the tangent is not influenced by the speed in the preceding and subsequent curves (Table 19).

Table 19. Minimum tangent length needed in order to achieve the maximum predicted speed for tangents (TLmax).

Formula for …	V85 in the curves [km/h]	$V85_{max}$ on the tangent [km/h]	TL_{max} [m]
Germany	96.68	120.92	478.71
Worldwide	88.28	105.31	299.35

Note. Based on Lamm et al. (2007) by using the formula for Germany and for worldwide application. $CCR_S = 258.79$ [gon/km] and acceleration = 0.85 [m/s^2].

The comparison of the TL_{max} values shown in Table 19 with the actual tangent lengths used in the simulation revealed that only the long experimental straight road sections T08, T09, and T16 were independent (see Appendix A 3.7). This was not the case for the tangents before experimental curves, at least not after the changes were made to the original plan (see above). Assuming that the formulas of Lamm et al. can be applied to the data collected in the simulator (see also chapter 4.3.3.13), this means that maximum speed on the tangent preceding an experimental curve was not only influenced by the experimental variation of the curve, but also by the speed in the preceding curve.

Because acceleration and deceleration values vary between participants and between curves, it is not possible to determine which part of the maximum speed on the tangent is due to the preceding section and which part is due to the experimental section. Similarly, variables for which the values are related to speed, such as the distance of deceleration from curve beginning, cannot be used. Therefore, it was decided to entirely discard speed and related parameters for the tangents preceding experimental curves.

Thus the only remaining driving behaviour parameter with which to assess the influence of the experimental variation is minimum speed in the curve itself. Is there a way to make sure that this parameter is not influenced by the speed in the preceding curve? A tentative answer to this question is provided by Lamm et al. (2007) themselves because they only use the geometry of the curve itself to predict speed for curves without taking into account speed on the preceding tangents.

A second, more relevant answer can be provided by using the actual V85 values based on the data collected in the simulator instead of the predicted, calculated V85 values based on Lamm et al. (2007). Actual V85 values were calculated based on the minimum speed values in the curves and the maximum speed values on tangents. V85 was used instead of average speed to ensure comparability with the Lamm et al. approach. According to the results reported in the previous paragraphs, the speed on tangents that precede experimental curves is likely to be influenced by speed in both the preceding as well as the subsequent curve. However, if minimum speed in the experimental curves is to be used for further data analysis, speed in the experimental curves should not be influenced by speed in the preceding curves.

To statistically test this assumption, the V85 values for the road elements were arranged in a matrix so that V85 in the curve and on the tangent, both preceding the experimental curve, and V85 in the experimental curve itself can be correlated. The values for the outbound direction were also included in this matrix, although they are not included in the final data analysis (see chapter 4.3.3.11). The reason for this is that the issue of insufficient familiarisation is not relevant for the question at hand because the laws of physics apply regardless of familiarization. Furthermore, this allowed the number of cases in the analysis to be doubled which is highly useful given that this matrix contains only five experimental curves for each direction.

To estimate the influence on measured V85, two linear regression analyses were performed following the rationale described above. The first regression analysis was performed for V85 on the tangent as dependent variable and V85 of both the preceding curve and the subsequent experimental curve as independent variables. The second regression analysis was performed with V85 in the experimental curve as dependent variable and with V85 of both the preceding tangent and the preceding curve as independent variables.

Table 20. Influence of V85 in the preceding curves and V85 in the experimental curves on V85 on the tangents preceding the experimental curves: results for a linear regression.

Variable	*B*	*SE B*	*β*	*t*	*p*
Constant	14.32	21.32		9.94	.00
V85 preceding curve	0.41	0.22	.40	1.86	.11
V85 experimental curve	0.59	0.20	.64	2.93	.02

Note. $R^2 = .69$, Corrected $R^2 = .60$, $F(2,7) = 7.84$, $p = .02$.

Table 21. Influence of V85 in the preceding curves and V85 on the preceding tangents on V85 in the experimental curves: results for the linear regression.

Variable	*B*	*SE B*	*β*	*t*	*p*
Constant	13.50	12.28		9.94	.00
V85 preceding tangent	0.94	0.32	.87	2.93	.02
V85 preceding curve	-0.26	0.33	-.24	-0.81	.45

Note. $R^2 = .58$, Corrected $R^2 = .46$, $F(2,7) = 4.79$, $p = .05$.

The results shown in Table 20 and Table 21 are in line with the assumptions: V85 in the experimental curves was not influenced by V85 in the preceding curves (Table 21). The fact that V85 on the tangent was more influenced by speed in the subsequent experimental curve than by speed in the preceding curve (Table 20) is regarded as an encouraging result and could even be interpreted in favour of an additional analysis of parameters on the tangents preceding the curves. However, the results in Table 20 and Table 21 are based on a very small sample. Therefore, it was conservatively decided to use minimum speed in the experimental curves as the only reliable parameter to assess the effect of the experimental variation.

4.3.3.13 Ecological Validity of the Driving Simulator Study

Ecological validity is that part of external validity which deals with the transferability of the results found in the experimental setup to real settings. With respect to driving simulators, ecological validity is achieved if data collected in the simulation can be mapped to data collected in real driving situations in the field. While, in principle, this question would have to be posed every time new environmental or driving dynamic conditions are simulated, it is often done just once for a sample of situations, due to the considerable effort and expenditure involved. Similarly, it would have to be done for every single simulator because the technical details and the software used to simulate driving dynamics differ considerably between simulators. Again, due to the effort and expenditure involved, such validation is often merely done by citing evidence found for comparable simulators (for example, fixed-base versus moving-base simulators).

The driving simulator used for the study in this thesis has not yet been externally validated. This validation is currently underway with the data collected on the intermittent section between the outbound and inbound direction of the simulator experiments reported here (see chapter 4.3.3.3). The data collected on this intermittent section in the simulator will be compared to data collected with the same participants during driving studies on the real B6. However, because the external validation of the simulator is only a side aspect of this thesis, the effort needed to analyse the data for both the simulated drives and the real-road drives could not be invested as part of this thesis.

Therefore, instead of a sophisticated validation which compares real world and simulated data, the current chapter reports a tentative validation that makes use of three different approaches:

- a theoretical validation by referring to existing validation studies reported in the literature;
- a combination of theoretical and empirical validation by applying the formulas of the Lamm et al. (2007) assessment described in the preceding chapter and comparing the results with the data collected in the simulator; and
- a comparison of group-specific differences in behaviour reported in the literature with the respective group-specific differences in the simulation.

With the increased availability of simulation tools, the question of ecological validity is gaining additional weight, both for companies selling this technique and even more so for researchers using it. In most studies which deal with this subject, a distinction is made between absolute and relative ecological validity (Godley, Triggs, & Fildes, 2002; Kaptein, Theeuwes, & van der Horst, 1996; Reimer, D'Ambrosio, Coughlin, Kafrissen, & Biederman, 2006). Relative validity means that the direction and order (i.e. the rank) of differences found between two or more conditions in the field is the same as in the simulator. Absolute validity also requires that the extent of the differences is the same or at least comparable.

With regard to behavioural data such as speed and lateral distance, it is usually found for simulator studies that relative validity is high, but absolute validity is low (Godley et al., 2002; H. C. Lee, Cameron, & Lee, 2003; Reed & Green, 1999; Reimer et al., 2006; Törnros, 1998). Other authors also report absolute validity with respect to speed, at least for some sections on rural roads (Bella, 2008), or even in general when approaching junctions (Yan et al., 2008). Similar results are reported in Schwebel, Gaines & Severson (2008) for the crossing behaviour of pedestrians, although the significant correlation between virtual reality and reality was quite low. Reimer et al. (2006) successfully applied a comparison between self-reported past behaviour and behaviour in the simulator to validate a simulator. Although one could assume that more sophisticated simulators increase ecological validity, such evidence is scarce or even contradictory. For example, Reed & Green (1999) found no differences concerning validity between high and low scene fidelity levels. Alm (cited in Kaptein et al., 1996) found that adding a moving-base increases validity of lateral data, but has no effect on the validity of speed data.

From this overview, no effects were identified which would indicate any problems regarding relative validity for the simulator used in this thesis. However, it must be added that even the most sophisticated simulators are not as yet able to induce realistic levels of demand as found in real driving (Carsten & Brookhuis, 2005). The same can be assumed for perceived risk or other relevant motivational factors as introduced in chapter 2.3. The relevance of this shortcoming was pointed out by Bella (2008) who held inappropriate risk perception responsible for the fact that absolute validity could not be found for all locations (see above).

A second proxy validation was performed by comparing V85 calculated according to the formula for Germany reported in Lamm et al. (2007) (see preceding chapter) with measured V85 found in the simulation (Figure 25). This data was tested for correlation and differences by using a *t*-test for independent samples (Table 22). The data was tested prior to the *t*-test for normal distribution with the K-S test. This test was not significant for measured V85 (*p* = .94), but was significant for calculated V85 (*p* = .04), indicating non-normally distributed data for this parameter. Because the t-test is robust against this precondition of normally distributed data (Brosius, 2008, p.

465) it was decided to use it nevertheless. To account for this violation, the results given in Table 22 are the ones for ‘equal variances not assumed’, which are more conservative (Brosius, 2008), despite the fact that Levene’s test indicated equal variances ($F = 2.13$; $p = .15$).

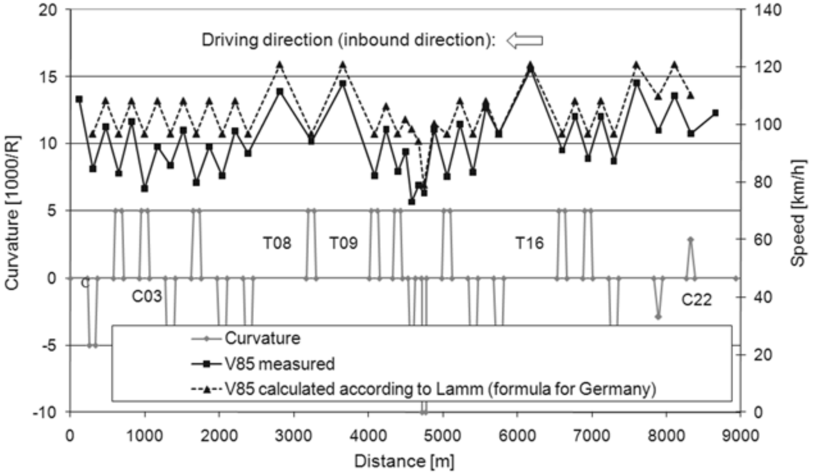


Figure 25. Element-wise comparison of measured V85 and calculated V85 (according to Lamm, 2007) for the inbound direction.

Table 22. Correlation (Pearson) and difference (*t*-test for independent samples) between calculated V85 according to Lamm et al. (2007) and measured V85 for all elements in the inbound direction.

Variable	*M*	*SD*	*r*	*p (r)*	*t*	*df*	*p*
V85 calculated	103.43	9.33					
V85 measured	93.39	11.51					
Difference	9.40	2.35	.89	.00	4.01	75.98	.00

Although drivers in general drove more slowly in the simulator in comparison to the calculated speed (as proxy for speed driven in the field), both speeds are significantly correlated (for both see Table 22).

Next, a linear regression was performed with measured V85 as the independent variable and calculated V85 according to Lamm et al. (2007) as the dependent variable (Table 23). Taken together, the data collected in the simulator showed low absolute, but high relative validity when compared to calculated V85 according to Lamm et al..

Table 23. Influence of V85 measured on V85 calculated according to Lamm et al. (2007): results for the linear regression.

Variable	*B*	*SE B*	*β*	*t*	*p*
Constant	36.23	5.78		6.27	.00
V85 measured	.72	.06	.89	11.71	.00

Note. R^2 = .79, Corrected R^2 = .78, $F(1,37) = 137.17$, $p < .01$.

Finally, a third tentative validation was conducted by analysing group-specific differences in behaviour. For this purpose, the demographic variable age was selected because its effect on driving behaviour is as well documented as its relevance to traffic safety (chapter 2.2). As was preliminarily shown for the data collected in the laboratory (chapter 4.2.5.7), it is also relevant with regard to the model. Based on the literature (chapter 2.2), a main effect of age or at least an interaction effect of age and task complexity on behaviour would indicate ecological validity of the simulation.

Task complexity was represented in the simulator study by road geometry, with curves representing complex situations and tangents representing simple situations. To statistically test the assumptions, minimum speed was averaged for all experimental curves (curves C02, C03, C05, C15, and C19) and maximum speed was averaged for all experimental tangents (T08, T09, T16). Because age could have interacted with familiarisation of driving in the simulator only the data for the inbound direction were selected, for which effects of familiarization were ruled out in previous analyses (chapter 4.3.3.11). Age was used both as covariate and as factor, with participants having been assigned to age groups according to the rules described in chapter 4.3.3.5. For the statistical analyses, univariate repeated-measures ANOVAs were used (see chapters 4.2.4.5 and 4.3.3.15 for details on the statistical methods).

The ANOVA with age as covariate resulted in a significant main effect, both for condition, $F(1, 41) = 17.29$, $p < .01$; $\eta^2 = .30$, and for age, $F(1, 41) = 4.99$, $p = .03$, $\eta^2 = .11$. In addition, the interaction effect of age and condition was significant, $F(1, 41) = 17.29$, $p < .01$, $\eta^2 = .30$. Similar results were obtained by using the four age groups as factor: a main effect of condition was found with $F(1, 39) = 421.53$, $p < .01$, $\eta^2 = .92$, together with a significant interaction effect of age group and condition, $F(1, 39) = 3.72$, $p = .02$,

$\eta^2 = .30$. However, for this analysis, the main effect of age group was not significant, $F(3, 39) = 1.97$, $p = .13$, $\eta^2 = .13$.

While the main effect of condition was of course due to the higher speed on the straight road sections, the effect of age was further tested statistically by calculating two univariate ANOVAs separately for curves and for tangents with age group as factor. In accordance with the interaction effect of age group and condition reported above, these two analyses resulted only in a significant main effect of age group for a single condition, namely the curves. Here, the oldest age group drove significantly slower ($M = 63.45$, $SD = 13.30$) than the youngest age group ($M = 76.85$, $SD = 6.80$) ($p = .02$). The differences to the other two age groups were not significant (Group II: $M = 68.57$, $SD = 9.66$; Group III: $M = 63.95$, $SD = 7.70$). The fact that the difference became significant despite the young age of the oldest age group (see chapter 4.3.3.5) is interpreted as supporting the age effect.

In general, the results replicated the findings reported in the literature and were thus a further indication of the ecological validity of the simulator study. Of additional interest, and in accordance with the literature, the variance of the values, which is reported above as standard deviation, increased with age and was nearly double the amount for the oldest age group in comparison to the youngest age group.

It is concluded that the three tentative validation approaches can all be interpreted in favour of the ecological validity of the simulator or at least the simulator study conducted for this thesis. Therefore, the results reported in the next chapters should also be found in real environments. Of course, until the validation is completed with the data collected on the real road, results should be interpreted with due caution.

4.3.3.14 The Dependent Variables: Final Version

In chapter 4.3.3.12 it was shown that the tangents before the experimental curves were not independent which means that data collected there is also influenced by the preceding road element. Because it is not possible to disentangle the influence of the preceding curve and the experimental curve on the tangent, it was decided to only use the minimum speed in the curve as the relevant dependent parameter. Minimum speed was assessed for the section covering 50 m before the curve (see also Figure 24) up to the curve ending. For the straight road sections, two parameters were selected:

- the maximum speed, assessed for the entire tangent length, and
- the average speed, assessed from 200 m after the beginning up to 200 m before the end of the straight road section to minimise the influence of the preceding and subsequent road element.

Because of the findings in chapter 4.3.3.11, only the inbound direction was analysed. Unfortunately, this decision also affected the use of the subjective data. While the curve direction of all experimental curves was right when driven in the inbound direction, the same curves were all shown as left curves in the videos used to elicit the subjective ratings. However, in my opinion, the ratings can still be compared because the difference in direction is consistent throughout all curves, with driven curves all being right curves and rated curves all being left curves.

4.3.3.15 Statistical Methods of Data Analysis

The rationale behind the simulator study was to test the effect of environmental design (the independent factor) on perception, expectations and behaviour (the dependent variables). 'Perception' was assessed by selected RECL items, expectations were assessed specifically by the item 'the road element requires reduced speed', and behaviour was assessed by the speed parameters described in the preceding chapter. Because the data for the different conditions were collected with the same participants, statistical methods for repeated measures or dependent samples were used.

In Hypothesis 1, it was assumed that speed is influenced by environmental characteristics. This assumption was tested by analysing differences between the environmental variations. The analysis of such differences was performed separately for straight road sections and curves because of the inherent differences between these two geometric elements. As both straight road elements and curves contained more than two variations, tests had to be chosen which accounted for this number of variations. The selection of a specific test also depended on the distribution of the data. Therefore, at first normal distribution of the data was tested separately within each experimental condition with the K-S test. If the data were not normally distributed (significant K-S test, see chapter 4.2.4.4), Friedman chi-square tests for several related samples were used. If the data were normally distributed, repeated-measures analyses of variance (ANOVAs) were used (see chapter 4.2.4.5). This first test of overall difference was conducted to account for the accumulation of alpha-errors when several *t*-tests are calculated (Bortz, 2005; Field, 2009; Rudolf & Müller, 2004).

The Hypotheses 1a to 1c assumed specific differences between one design variant in comparison to a reference design or other design variants. Statistically, the analysis of overall differences described above can be seen as a prerequisite for the subsequent pair-wise comparisons to be conducted as part of the specific Hypotheses 1a to 1c. If the data were not normally distributed, Friedman's chi-square test for several related samples used in the first step (see above) was followed by non-parametric Wilcoxon tests for paired samples. If normal distribution could be assumed and GLMs were used, two possibilities are implied in the GLM-model in SPSS for analysing differences between two conditions: either the use of contrasts or the post-hoc use of *t*-tests (for paired samples). However, for the data at hand, none of these possibilities was used in the first place. The reason is that contrasts, like the Helmert-contrast, test the difference between one condition and the averaged values of all other conditions (Field, 2009). While such a comparison might be helpful in some cases, it was not the main interest as formulated in the hypotheses and was therefore discarded.

Furthermore, instead of using the pair-wise comparisons implemented in the GLM-model, additional *t*-tests for paired samples were calculated separately. This was done because in the SPSS-output the correlation and its significance are given in addition to the *t*-test results. The former is regarded as especially useful for the interpretation of driving experiments because it allows an estimation of intra-individual stability of behaviour across conditions. Nevertheless, in order to account for the alpha-error accumulation in several separated tests, the results for the pair-wise comparison in the GLM were also used. The latter were calculated with the Bonferroni correction and thus are adjusted for alpha-error accumulation. The Bonferroni correction is regarded as the most robust technique (Field, 2009). If the results of both tests differed, both are given in the text; otherwise, only the separated *t*-test results and the correlations are given.

Hypothesis 2 assumed that differences in behaviour were mirrored by differences in the subjective ratings. Testing this hypothesis required, in a first step, conducting the same analyses for the subjective ratings as were conducted for behaviour to test Hypothesis 1 (see above). The selection of a specific statistical method depended again on the analyses of its assumptions as described above. Only after differences in the subjective ratings between different road designs were assessed based on the ratings of each participant could the commonalities between behaviour and subjective ratings then be analysed.

The analysis of these commonalities is essentially a question of correlation, the statistical background of which was explained in chapters 4.2.4.5 and 4.2.4.6. For this analysis of correlation, the data were averaged across each design element, separately for behaviour and the subjective ratings. The data basis for the analysis of correlation was thus reduced to the number of experimental road elements. Usually, the choice of parametric or non-parametric correlation coefficient is based on whether the data are normally distributed or not. However, the K-S test usually used to test this assumption is not powerful for small samples (Rudolf & Müller, 2004). Therefore, even if the assumption of normally distributed data could not be discarded as a result of the K-S test, Kendall’s non-parametric rank correlation was calculated in addition to Pearson’s coefficient.

In contrast to the curves, performing a correlation analysis was not appropriate for the straight road sections due to their small number. For the three straight road sections a descriptive in-depth analysis of the commonalities and differences between the results for behaviour and the subjective ratings is more promising than calculating a numerical measure of correlation alone. Questions relating to this issue are further discussed in the last paragraph of this chapter.

The additional research question addressed group-specific differences between and within the conditions analysed in Hypotheses 1 and 2. A group-specific factor could have been included in the repeated-measures ANOVAs as part of the analyses for Hypotheses 1 and 2 because the same statistical methods were used. However, it was decided to perform separate analyses with group-specific factors in order to account for the limited knowledge regarding differentiated effects of driver group and environment. If the data were normally distributed, the individual factors as described in chapter 4.3.3.5 were included in the ANOVAs as between-subjects factors. Separate repeated-measures ANOVAs were calculated for each factor because of the small sample size and interdependencies between the different groups when grouped according to age, gender, or driving style (see chapter 4.3.3.5).

If the overall results for the between-subjects factor were significant, additional univariate analyses of variance were calculated separately for each condition (i.e. a specific curve and a specific dependent variable) and the respective between-subjects factor. If the between-subject factor contained more than two groups, differences between the groups were analysed with post-hoc Scheffeé-tests, implemented in the GLM. According to Bortz (2005), this test is robust against violated assumptions and tends to be conservative. If the data were not normally distributed, and subsequently ANOVAs could not be used, non-parametric tests were chosen, depending on the number of groups in the sample. For two independent sub-groups the Mann-Whitney test was used; for more than two independent sub-groups the Kruskal-Wallis test was used.

At the end of this chapter one question remains open: why were the model assumptions tested separately according to the hypotheses instead of testing the model assumptions as a whole within a single statistical procedure such as structural equation models (SEM, see chapter 4.2.5.7)? The reason is that such a procedure was not possible with the data. This is even independent of the additional model assumptions with regard to safety which were analysed in the field study. Of course, at first sight, it would have been possible to analyse the relationship between perception, expectation and behaviour. However, with the available dependent data, this analysis would have been restricted to a single road condition and would have had to be calculated anew and separately for each other road condition. Influences caused by different road and environmental conditions as proposed by the model could thus not have been assessed.

A potential solution would have been to restructure the data so as to change the dependent data structure into a pseudo-independent one. Such an approach was described for the factor analysis conducted with the RECL items in the laboratory task (see chapter 4.2.5.1). However, besides the violation of the assumptions for SEM, this would still not have allowed isolation of the effect caused by road design. In order to do this, differences between road designs would again have had to be tested in the way described above. Testing of the model assumptions as a whole within a single statistical procedure would therefore require a large sample of different road designs, not necessarily participants. The averaged data across all participants for each road condition would then be used to test the model assumptions. This would of course require considerable effort, especially concerning data collection.

4.3.4 Results

4.3.4.1 Speed on Straight Road Sections: Hypotheses 1a and 1b

In this chapter, differences in speed between the three long straight road sections were analysed for the inbound direction. Because these straight road sections only differed in terms of their environmental design (monotonous, diversified and tree-lined) and not in terms of length and cross-section, differences in behaviour can be attributed to these environmental variations.

As described above, two relevant parameters were chosen for the variable speed: the maximum speed of each driver assessed for the entire straight road section and the average speed throughout the middle section of the straight road section. Figure 26 gives an overview of the data and their development across the entire straight road sections.

As the data for both the maximum and the average speed were normally distributed, ANOVAs for repeated measures were calculated, followed by separate *t*-tests for paired samples. The ANOVAs indicated highly significant differences

- for maximum speed, $F(2, 41) = 5.97$, $p = .01$, $\eta^2 = .23$, and
- for average speed, $F(2, 41) = 5.58$, $p = .01$, $\eta^2 = .21$.

The results for the subsequent *t*-tests for paired samples are shown in Table 24 for maximum speed and in Table 25 for average speed, together with the correlations. The significant difference in maximum speed between the diversified and the tree-lined road (Table 24) was no longer significant when calculated with Bonferroni's correction ($p = .16$). No such effect was found for average speed (p with Bonferroni's correction $= .03$).

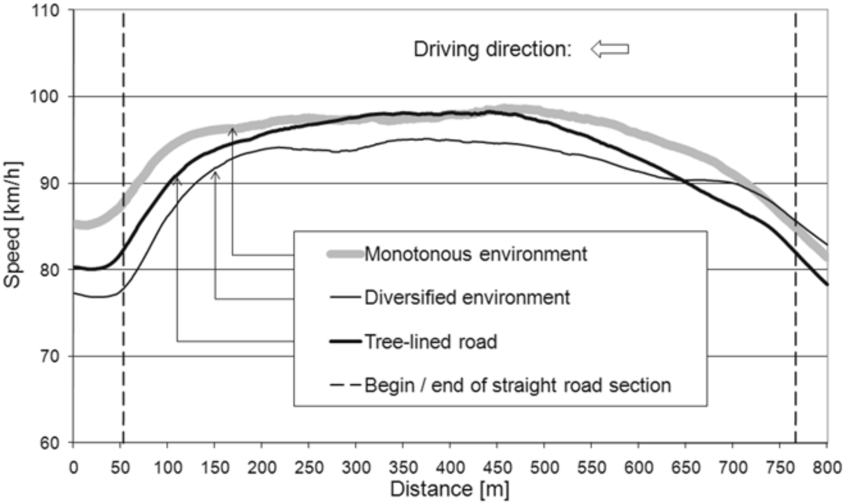


Figure 26. Speed [km/h] averaged across participants for the inbound direction on the long straight road sections.

Table 24. Maximum speed on straight road sections for the inbound direction: results of the *t*-tests for paired samples.

Design	*M*	*SD*	*r*	*p (r)*	*t (df* = 42)	*p*
Diversified (T08)	99.11	11.83				
Tree-lined (T09)	101.48	12.51	.80	.00	-2.00	.05
Diversified (T08)	99.11	11.83				
Monotonous (T16)	102.97	14.84	.87	.00	-3.40	.00
Tree-lined (T09)	101.48	12.51				
Monotonous (T16)	102.97	14.84	.82	.00	-1.14	.26

Table 25. Average speed on straight road sections (middle section; inbound direction): results of the *t*-tests for paired samples.

Design	*M*	*SD*	*r*	*p (r)*	*t (df* = 42)	*p*
Diversified (T08)	94.24	10.79				
Tree-lined (T09)	97.34	11.72	.79	.00	-2.74	.00
Diversified (T08)	94.24	10.79				
Monotonous (T16)	97.82	14.25	.85	.00	-3.08	.00
Tree-lined (T09)	97.34	11.72				
Monotonous (T16)	97.82	14.25	.85	.00	-0.42	.68

The results indicate that environmental variation had a strong influence on speed. Diversified environments reduced speed and monotonous environments induced faster speeds. It is interesting to note that a tree-lined road had the same effect as a monotonous environment void of all objects. The highly significant correlations in all tests also indicate high intra-individual stability in behaviour. Speed was therefore further analysed with respect to the additional research question concerning differences between driver groups.

4.3.4.2 Differences in Speed Between Driver Groups on Straight Road Sections

The analysis of individual and group-specific differences on straight road sections is supported by the results shown in Table 24 and Table 25. The significance of the correlations shown there indicates that behaviour is stable for individual drivers between different elements. This in turn can be seen as a prerequisite for significant main effects of driver group on behaviour. In order to also account for potential interaction effects between environment (i.e. the experimental variation on the straight road sections) and driver group, ANOVAs were calculated with environment as factor and driver group as between-subjects factor. The ANOVAs were only calculated for average speed because this parameter resulted in more pronounced differences between the straight road sections (see Table 24 versus Table 25 in the preceding chapter). Changes in the significance of the between-subjects factors caused by adding the within-participants factors were only reported if they fundamentally changed the interpretation reported in the preceding chapter (i.e. from significant to non-significant or vice versa). Results for the ANOVAs and the within-subjects factors were as follows:

- age group, $F(3, 39) = .87$, $p = .46$; environment × age group, $F(6, 78) = .99$, $p = .44$;
- gender, $F(1, 41) = 2.97$, $p = .09$; environment × gender, $F(2, 40) = 2.85$, $p = .07$;
- item 3 'I like driving fast on straight rural roads', $F(1, 41) = 2.46$, $p = .12$; environment × item 3, $F(2, 40) = .21$, $p = .81$;
- item 7 'Compared to other drivers I tend to drive slower on rural roads', $F(1, 41) = 2.27$, $p = .14$; environment × item 7, $F(2, 40) = .72$, $p = .50$; and
- item 9 'How would you describe your driving style?', $F(1, 41) = .44$, $p = .51$; environment × item 9, $F(2, 40) = 1.13$, $p = .33$.

Summing up the results it can be stated that none of the variables named above showed a significant effect, neither as a main effect nor as an interaction effect with the experimental conditions.

The non-significant main effects for all driver groups are somewhat surprising given it was initially considered that the high degree of freedom on straight road sections stresses group-specific differences in behaviour. Such differences are usually reported in the literature (see chapter 2.2) and were also found in the laboratory study, albeit for rated speed (see chapter 4.2.5.7). Given that such effects were in fact found in reality, two potential reasons could account for the differences in the results found in the simulator:

- either a selection effect relating to the participant sample, or
- a ceiling effect of speed in the simulator.

The latter effect is unlikely because differences in speed were found between the different environmental conditions (see preceding chapter). In contrast, the former effect almost certainly played a role as already described in chapter 4.3.3.5 with the young age of the oldest age group.

Strictly speaking, the non-significant results did not indicate a need for further analyses. However, because these non-significant results were probably influenced to a large extent by sample characteristics, it was decided to at least further analyse the differences between those groups for which the effects were closest to significance. This was the case for the factor gender, for which both the main effect and also the interaction effect approached significance. These effects were further analysed with univariate analyses of variance, separated by condition. The descriptive statistics for the two groups and the different environmental conditions are shown in Figure 27.

The difference between male and female drivers was significant for the tree-lined road, $F(1, 41) = 4.53$, $p = .03$, $\eta^2 = .10$, but not for the monotonous environment, $F(1, 41) = 3.21$, $p = .08$, $\eta^2 = .07$, nor for the diversified environment, $F(1, 41) = .77$, $p = .38$, $\eta^2 = .02$. The fact that the effect for the monotonous environment merely approached significance might again be attributable to the unbalanced sample characteristics. In this case the effects could be summarised as follows: the diversification of the environment clearly had a positive effect when compared to the other two conditions, with male drivers reducing speed and female drivers not being negatively affected. This resulted in a general decline in average speed and a reduction in differences between driver groups, that is, an increase in the homogeneity of speed. Both effects will have a positive effect on road safety. Whether these effects and the effects found in the preceding chapter are mirrored by the rated road characteristics is the subject of the next chapters.

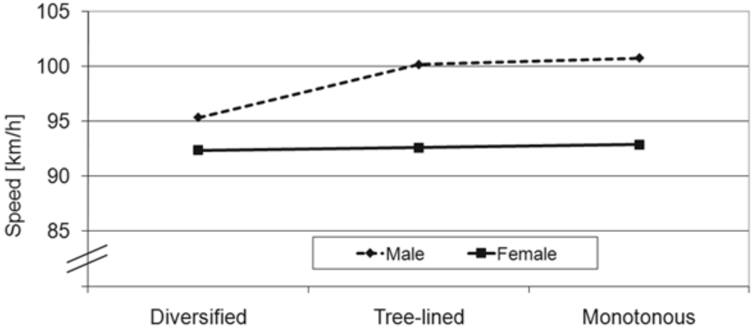


Figure 27. Differences between male and female drivers in average speed for the middle section of the long straight road sections.

4.3.4.3 Perception and Expectations on Straight Road Sections: Hypotheses 2a

The central question dealt with in this chapter is whether the objective data on the long straight road sections is mirrored by subjective ratings. In chapter 4.2.5.5, the two items 'dangerous' and 'monotonous' were selected as marker items to assess the perception of behaviourally relevant road characteristics. Their selection was based on the analysis of the RECL structure and on theoretical considerations.

In Figure 28 these two marker items are depicted together with additional items for the long straight road sections in the simulator study. The item 'demanding' was additionally selected to account for the theoretical debate on whether demand and subjective feeling of risk can be distinguished (Fuller, 2005; Fuller et al., 2008). In addition, the item 'The road element requires reduced speed' (abbreviated as 'reduced speed') was selected to assess expectations and behavioural intentions. Finally, the item 'enjoyable' was selected to account for hedonic quality (Berlyne, 1970; Steyvers, 1993). The data shown in Figure 28 were tested statistically for differences (Table 26). Non-parametric tests were used because all item values differed significantly from the assumption of normal distribution.

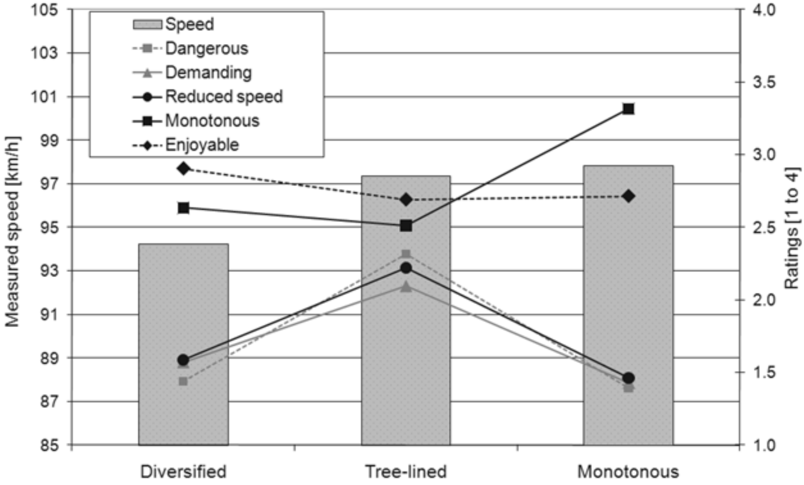


Figure 28. Measured speed (M for middle section) and subjective ratings for the long straight road sections.

Table 26. Subjective ratings for long straight road sections: descriptive statistics and statistics for Friedman's post hoc tests.

	Diversified	Tree-lined	Monotonous		
Variable	*M (SD)*	*M (SD)*	*M (SD)*	χ^2 (*df* = 2)	*p*
Dangerous	1.44 (0.55)	2.32 (0.79)	1.39 (0.59)	39.46	.00
Demanding	1.57 (0.59)	2.10 (0.73)	1.43 (0.59)	21.68	.00
Monotonous	2.63 (0.83)	2.51 (0.84)	3.32 (0.82)	26.57	.00
Enjoyable	2.90 (0.69)	2.69 (0.75)	2.71 (0.71)	3.21	.20
Reduced speed	1.59 (0.55)	2.22 (0.65)	1.46 (0.50)	38.64	.00

Except for the item 'enjoyable', all items resulted in highly significant differences which were tested pair-wise for each pair of environmental conditions with non-parametric Wilcoxon tests (Table 27, results for 'enjoyable' are given in brackets).

Table 27. Subjective ratings for long straight road sections: statistics for the non-parametric Wilcoxon test for paired samples.

	Tree-lined - Diversified		Monotonous - Diversified		Monotonous - Tree-lined	
Variable	*Z*	*p*	*Z*	*p*	*Z*	*p*
Dangerous	-4.58	.00	-0.58	.77	-4.48	.00
Demanding	-3.39	.00	-1.50	.21	-3.85	.00
Monotonous	-0.65	.56	-3.57	.00	-4.03	.00
(Enjoyable	-1.29	.22	-1.55	.17	-0.12	.98)
Reduced speed	-4.25	.00	-1.29	.31	-4.63	.00

The results stress the exceptional effect of tree-lined roads. They were rated as being more dangerous, demanding and monotonous as well as requiring a higher reduction in speed than the other two conditions. In contrast to these ratings, driven speed was faster for the tree-lined roads, at least in comparison to the diversified road (see chapter 4.3.4.1). Thus, the ratings at least for the tree-lined road did not mirror behaviour. Differences in the ratings between the diversified and the monotonous environment were only found for the item 'monotonous' which was also rated as being more monotonous. This item was in line with Hypothesis 2a and mirrored the faster speeds which were found on the monotonous road.

4.3.4.4 Differences in Perception Between Driver Groups for Straight Road Sections

The ratings representing perceived road characteristics were further analysed for differences between driver groups. In line with chapter 4.3.4.2, only the difference between male and female drivers was tested.

The underlying research question was again, whether the differences in speed for these two groups were mirrored by the rated characteristics of the road environments. Mann-Whitney tests, which were calculated with the grouping variable 'gender' for all items and all straight road sections, resulted in only one single significant difference between male and female participants: this was for the item 'dangerous', with the tree-lined road being rated as less dangerous by the female drivers (Table 28).

Table 28. Differences between female and male participants in the ratings of the item 'dangerous' for the long straight road sections (Mann-Whitney test).

	Females			Males				
Design	*M*	*M* Rank	Sum Rank	*M*	*M* Rank	Sum Rank	*Z*	*p*
Diversified	1.44	20.5	328.5	1.42	21.3	532.5	-0.23	.86
Tree-lined	2.00	16.1	257.5	2.52	24.1	603.5	-2.29	.02
Monotonous	1.44	21.8	348.0	1.35	20.5	513.0	-0.39	.77

The results further stress the exceptional situation of tree-lined roads: differences between the two participant groups are only present for this environment. However, the direction of the difference is in conflict with the findings for speed (chapter 4.3.4.2). Thus, when analysed on a group-specific level, behaviour on the straight road sections is not mirrored by the perception of the road characteristics, at least not for the long straight road sections.

4.3.4.5 Summary and Discussion of Results for the Straight Road Sections

With respect to road safety, the most important result from the preceding chapters is that behaviour on straight rural roads is influenced by environmental characteristics. That is, in line with the assumptions made in *Hypothesis 1a*, speed was reduced in a diversified environment compared to a tree-lined or a monotonous environment. In line with the assumptions of *Hypothesis 1b*, speed was faster on the tree-lined road compared to the other conditions. Concerning the behavioural effects of environmental variation with respect to road safety, it can be concluded that enriching the environment with objects will result in a decrease in speed. However, in order to decrease speed, these objects should be placed irregularly along the roadside to avoid them being perceived as single wall-like element.

In *Hypothesis 2*, it was assumed that the ratings which represented the perceived road situation and the expectations mirror behaviour. The findings only partly supported this hypothesis with only the item 'monotonous' showing differences which mirrored differences in behaviour. This is encouraging in so far as it supports the decision to select a second marker item from the perceptual factor in addition to the item from the motivational factor (see chapter 4.2.5.5). However, even the role of the item 'monotonous' as a possible explanation of the differences in behaviour was restricted to the differences between the monotonous road and the diversified road, and was not applicable for the tree-lined road. For the tree-lined road, the ratings even showed opposite effects in comparison to behaviour. This finding supports the findings of other authors (Ellinghaus & Steinbrecher, 2003; Zwielich et al., 2001) who point out the exceptional situation for tree-lined roads.

It is thus concluded that on straight rural road sections, behaviour tends more to be influenced subconsciously, for example, by the optic flow as described by Gibson (1986). The effects usually attributed to optic flow could, in contrast to the ratings, account for all the effects found for speed on the straight roads. This also applies to the tree-lined road: the trees would be perceived as a single, wall-like element for which optic flow is reduced in a similar way to a barren, monotonous environment. In addition, a pull-effect exerted on the drivers by the 'light at the end of the tunnel' probably strengthened the effect of the reduced optic flow.

A further impressive indication of the independence of ratings and behaviour on straight roads was found when analysing the *additional research question* concerning group-specific differences. Here it was found that males drove faster than females on tree-lined roads and to a lesser extent also on monotonous roads. However, it was also the male drivers who rated the tree-lined road as being more dangerous compared to the ratings of the female drivers. This result supports the assumption that driving on straight roads is far more influenced by non-conscious parameters such as the optic flow. Differences in the ratings can be attributable to differences in the awareness of the dangers of tree-lined roads. From a traffic-safety perspective (e.g. road safety campaigns in the media), it is interesting that this awareness did not affect behaviour.

Except for the differences between male and female drivers, no differences were found between driver groups on straight road sections, neither for behaviour, nor for perception. Two factors are seen as responsible for these results. Firstly, the characteristics of the sample may have influenced the results. Subjects were not preselected because inter-group differences were not the focus of this thesis. This resulted in the young oldest age group. Secondly, the results can be attributed to the low demand of long straight road sections. Differences might be more pronounced in more demanding situations such as curves. In fact, such interaction effects of demand and age are often reported in the literature (Weller & Geertsema, 2008).

4.3.4.6 Differences in Speed Between the Curves: Hypothesis 1c

The relevant speed parameter for curves is the minimum speed, calculated for the distance covering the entire curve and an additional 50 m before the curve (see chapter 4.3.3.14). The average values for all drivers are shown for the experimental curves and the reference curve in Figure 29. In this figure, the driving direction is from the right to the left in order to account for the inbound direction.

The parameter 'minimum speed' was normally distributed for all curves. The overall differences between the curves were highly significant as shown by the GLM for repeated measures, $F(4, 39) = 16.01$, $p < .00$; $\eta^2 = .62$. The results for the subsequent pair-wise comparisons of the experimental curves with the reference curve are shown in Table 29.

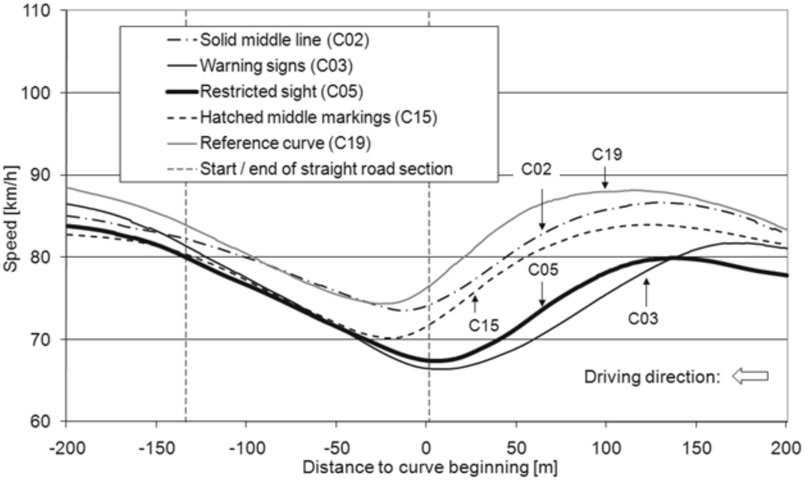


Figure 29. Speed [km/h] averaged across participants in the experimental and reference curves.

Table 29. Differences between the reference curve C19 and the experimental curves: minimum speed in the curves.

Curve design	M	SD	r	p (r)	t (df = 42)	p
Reference curve (C19)	72.76	14.74				
Solid line (C02)	72.07	11.45	.73	.00	0.44	.66
Signs (C03)	63.65	10.45	.48	.00	4.47	.00
Restricted sight (C05)	64.75	12.49	.76	.00	5.44	.00
Hatched markings (C15)	69.00	12.03	.73	.00	2.44	.02

The difference between the reference curve and the curve with hatched markings was no longer significant when calculated with the Bonferroni correction ($p = .19$). An additional t-test between the curve with signs and the curve with restricted sight resulted in no significant differences, $T = .84$; $p = .41$. Drivers showed highly consistent individual behaviour which is indicated by the significant correlations (see also Table 29). The correlation coefficient for the curve with signs is remarkably lower than for the other curves, but nevertheless highly significant.

All in all, the results show that single environmental variations that act as cues have an effect on behaviour. The cues caused a reduction in speed in comparison to the reference curve without cues. With regard to the intensity of the effect, formal signs showed the strongest effect together with informal restricted sight due to bushes planted on the inside of the curve.

4.3.4.7 Differences in Speed Between Driver Groups in Curves

Similar to the long straight road sections, differences in behaviour between driver groups were analysed for the curves. The first indication that such differences exist was given by the significant correlations shown in Table 29. As for the straight road sections, multivariate repeated-measures ANOVAs were performed separately for each between-subjects factor. The between-subjects factors used were the same as those for the straight road sections. The only exception was that grouping according to the values of the item 'I like driving fast on straight rural roads' was replaced by grouping according to the item 'I like driving fast on curved rural roads'. The conditions of the repeated-measures within-participants factor were all four experimental curves, together with the reference curve. The dependent variable used to characterise the curves was 'minimum speed in the curve'. The results were as follows:

- age group, $F(3, 39) = 4.95$, $p = .01$; design × age group, $F(12, 114) = 1.11$, $p = .36$; post-hoc Scheffé tests revealed two significant differences between the youngest age group I and the age groups III ($p = .02$) and IV ($p = .02$), with the youngest age group driving faster;
- gender, $F(1, 41) = .49$, $p = .49$; design × gender, $F(4, 38) = .34$, $p = .85$;
- item 4 'I like driving fast on curved rural roads': $F(1, 39) = .79$, $p = .38$; design × item 4, $F(4; 36) = .88$, $p = .48$;
- item 7 'Compared to other drivers I tend to drive slower on rural roads': $F(1, 41) = .12$, $p = .66$; design × item 7, $F(4, 38) = .82$, $p = .52$; and
- item 9 'How would you describe your driving style?', $F(1, 41) = .31$, $p = .58$; design × style, $F(4, 38) = .83$, $p = .52$.

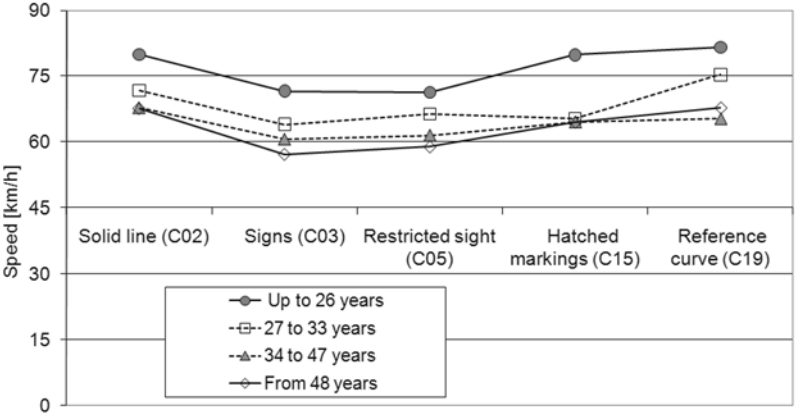


Figure 30. Differences in minimum speed in curves between the four different age groups.

Similar to when on long straight road sections, differences between driver groups were rare. For the curves they were only significant for the factor ‘age group’. These differences are depicted for each curve in Figure 30. The graphs shown in Figure 30 and the non-significant interaction effect of age group and curve design indicate high stability and consistency of the age effect. Despite this non-significant interaction effect, curve-specific differences between age groups were further analysed. This was done exploratory but might nevertheless have high relevance for an age-specific curve design.

For this purpose, univariate ANOVAs were conducted separately for each curve and with age group as between-subjects factor. Differences between the age groups were analysed with post hoc Scheffé tests in the case of equal variances. This was not the case for the curve with the solid line (significant results for Levene’s test, $F(3, 39) = 3.98$, $p = .01$). Therefore, for this curve, the results for Tamhane’s test are given instead. In Table 30, only the most relevant comparisons between the youngest age group I and the other age groups are given.

Interestingly, the differences between the youngest and older age groups were accentuated most for the curve with the hatched marking and the curve with the signs. They were not at all significant for the curve with restricted sight, although they were still present as can be seen in Figure 30. It should also be noted that this age effect starts to get significant for age group III, which has a lower limit of only 34 years. Differences between the oldest age group and younger age groups, apart from the youngest one, were not significant.

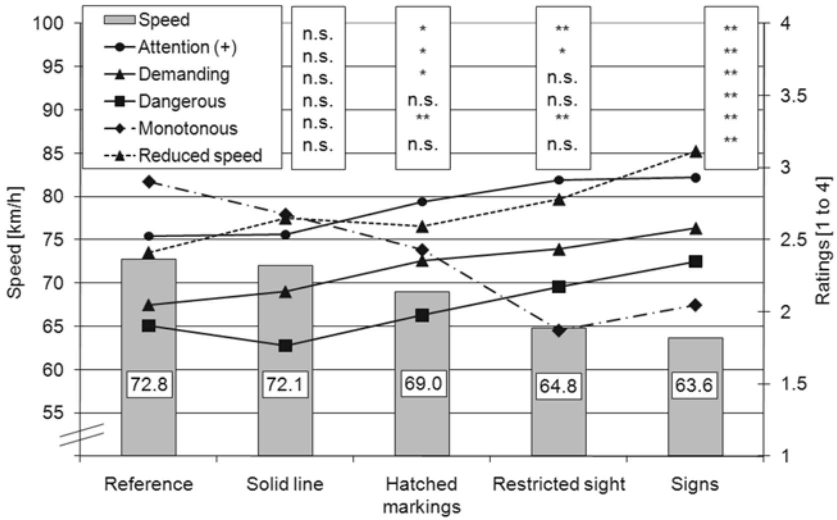
Table 30. Curve-wise differences in speed between age groups.

	ANOVA (corrected model)			Post hoc Scheffé test			
			Partial		Difference		
	F(3, 39)	*p*	η^2	Group I vs.	*M*	*SE*	*p*
Reference curve (C19)	3.35	.03	.21	Group II	6.29	5.84	.76
				Group III	16.31	5.69	.06
				Group IV	13.87	5.84	.15
Solid line (C02)	3.40	.03	.21	Group II	8.24	3.46	.16
				Group III	12.12	3.64	.02
				Group IV	12.33	5.28	.20
Signs (C03)	5.00	.01	.28	Group II	7.58	3.95	.31
				Group III	11.01	3.85	.06
				Group IV	14.38	3.95	.01
Restricted sight (C05)	2.34	.09	.15	Group II	4.84	5.11	.83
				Group III	9.78	4.98	.29
				Group IV	12.37	5.11	.14
Hatched markings (C15)	6.18	.00	.32	Group II	14.46	4.40	.02
				Group III	15.27	4.29	.01
				Group IV	15.34	4.40	.01

4.3.4.8 Perception of Curves and Expectations for Curves: Hypothesis 2b

Whether the differences in speed between curves were mirrored by differences in perceived curve characteristics and by behavioural expectations was analysed as done for the straight road sections. Because the ratings were not normally distributed, the Friedman rank test for several related measures was used item-wise to assess overall differences, followed by Wilcoxon tests to assess differences between the reference curve and each experimental curve.

The results of these Wilcoxon tests for selected items, which were all characterised by significant Friedman tests, are summarised in Figure 31. The detailed statistical results for all RECL-items and additional items are shown in Appendix A 3.9. The results indicate that the perceived road characteristics and the behavioural expectations of curves with additional cues differed from the reference curve without such cues. The direction of the results was further in line with the assumptions: the experimental curves were usually rated as being more dangerous, more demanding, and as being more monotonous than the reference curve. Furthermore, it was expected that they require more attention and a higher reduction in speed.



Note. Wilcoxon test results. *$p < .05$. **$p < .01$. n.s. = not significant.

Figure 31. Differences in measured minimum speed and subjective ratings for the curves.

Exceptions mainly concerned the curve with restricted sight. For this curve, the differences to the reference curve in terms of demand, danger and the expected required speed reduction were not significant (see Appendix A 3.9). This is surprising when the results for the adjacent curves with hatched markings on one side and with signs on the other side (Figure 31) are taken into account. These were significant with similar averaged values and similar standard deviation (see also Appendix A 3.9). Although the results for the non-parametric tests are based on ranks, not average and standard deviation, these parameters could be used as an indication that these non-significant differences are attributable to the reduced sample size for this curve (see chapter 4.3.3.9). However, in chapter 4.3.3.9 it was also shown that the subsample of participants with data for the curve with restricted sight did not differ from the subsample of participants without data for this curve with respect to relevant factors.

It can therefore be assumed that the results would have probably also been significant for this curve if the sample size had been the same as for the other curves.

Table 31. Correlation of speed in the curves with ratings of perception and expectation.

	Pearson		Kendall	
Speed with	*r*	*p*	*τ-b*	*p*
Attention	-.99	.00	-1.00	.00
Demand	-.97	.01	-1.00	.00
Danger	-.94	.02	-.80	.05
Monotony	.96	.01	.80	.05
Speed reduction#	-.87	.06	-.80	.05

Note. # Expected required speed reduction in the curve ahead.

The averaged values for each curve were used to test whether objective behaviour was mirrored by perception and expectations. Table 31 shows the results for the parametric and non-parametric correlation analysis. The very high and significant correlation coefficients shown in Table 31 support the assumption formulated in Hypothesis 2: perception and expectation mirror objective behaviour. Within the limits of the statistical method used (see chapter 4.3.3.15), these results are a preliminary but strong indication of the validity of the general model assumptions of the driving and driver behaviour model for rural roads.

4.3.4.9 Differences in Perception Between Driver Groups in Curves

As was done for the straight road sections, the ratings for the curves were analysed with respect to group-specific differences. In chapter 4.3.2, it was assumed that potential differences in behaviour between driver groups are also found for the perceived road characteristics. Besides the marker items, the tests were also carried out for the other RECL items and additional items. Using all items can be seen as additional indirect testing regarding the selection of the marker items (see chapter 4.2.5.5).

The selection of the grouping variable was based on the results for the objective data. For the curves, differences in speed were found between age groups (see chapter 4.3.4.7), which were thus also used for analysing differences between ratings. Overall differences between all four age groups were firstly analysed with the non-parametric Kruskal-Wallis test. If this test was significant, differences were further examined for this item and this curve. Here, Mann-Whitney tests were used to test differences between the youngest age group I and each of the other three age groups. The results of the Kruskal-Wallis tests (see Appendix, Table A 5) indicated significant or near-significant differences between age groups for only 10 out of 85 comparisons. Group-wise significant differences of these comparisons are shown in Table 32.

Table 32. Differences between age groups for the subjective ratings of curves: results for the Mann-Whitney tests.

	Group I		2nd Group			
Item	*M (SD)*	2nd G.	*M (SD)*	MW-*U*	*Z*	*p*
Reference curve						
Attention (+)	2.82 (0.40)	II	2.30 (0.48)	26.5	-2.34	.03
Attention (+)	2.82 (0.40)	IV	2.30 (0.48)	26.5	-2.34	.03
Concentration (-)	2.00 (0.45)	II	2.60 (0.52)	25.0	-2.48	.02
Wakefulness (-)	1.82 (0.40)	II	2.50 (0.53)	22.5	-2.75	.01
Wakefulness (-)	1.82 (0.40)	III	2.36 (0.50)	31.5	-2.44	.03
Solid line						
Demanding	1.83 (0.58)	II	2.50 (0.53)	27.5	-2.44	.02
Spacious	3.42 (0.51)	IV	2.50 (0.53)	17.5	-3.10	.00
Hatched markings						
Boring	1.45 (0.52)	II	2.10 (0.74)	28.5	-2.04	.06
Boring	1.45 (0.52)	III	2.09 (0.30)	25.0	-2.86	.01
Restricted sight						
Spacious	2.25 (0.46)	IV	3.20 (0.45)	4.0	-2.61	.02
Signs						
Attention (+)	3.00 (0.60)	IV	2.40 (0.70)	34.0	-1.95	.07
Concentration (-)	1.83 (0.58)	IV	2.60 (0.70)	27.0	-2.46	.02
Alertness (+)	3.08 (0.67)	IV	2.50 (0.53)	32.5	-2.02	.05

Note. 2nd G. = second age group; MW-*U* = Mann-Whitney-*U*;
(+) = increases [item name]; (-) = decreases [item name].

Besides the very few comparisons which showed any differences between the age groups at all, it is of interest that the differences do not all point in the direction one might expect. If motivational driving behaviour theories were applied to items such as 'The curve increases attention', one would expect that drivers who agree to this item more than other drivers would also drive more slowly. However, this was not the case. Both for the reference curve and the curve with signs, the age group which agreed more to the item 'The curve increases attention' (see Table 32) drove faster (see Table 30). Furthermore, the item 'The curve requires reduced speed', which in the model represents 'Expectations concerning appropriate future behaviour', did not show any significant differences between the age groups. The reason for this particular finding is probably the scale used to collect the ratings: this scale had only four points and thus would have required very pronounced driver group effects to result in significant differences in addition to the ones found for the experimental elements.

Despite this methodological limitation, the results showed that differences in behaviour between driver groups in the same curve cannot be attributed to differences in the perceived curve characteristics or to differences in the expectations concerning appropriate future behaviour.

4.3.4.10 Summary and Discussion of Results for the Curves

In *Hypothesis 1c* it was assumed that adding warning cues to curves would reduce speed. This hypothesis was supported by the data: speed was significantly reduced when cues were added to a reference curve without such cues. However, from a traffic-safety perspective, it is interesting to note that the reduction effect in speed was not only achieved by formal warning cues such as curve warning signs, but also indirectly by restricting sight in the curve. This means that informal cues could replace signs at selected locations, depending, for example, on rules and legislation. This would prevent an inflationary use of signs and at the same time ensure that signs were taken seriously by drivers in those situations where they are really needed.

In *Hypothesis 2b* it was assumed that changing curve characteristics by adding such cues would also change the perceived characteristics of these curves, together with the perceived appropriate behaviour for the curve. Such effects were found and mirrored the differences found for speed: curves which were perceived as more dangerous or demanding were also characterised by lower speeds. This finding indicates the importance of open-loop control in curve driving. Furthermore, the relevant input variables for this open-loop control might indeed be risk and expected workload as assumed by different motivational models of driving (see chapter 2.3).

In an *additional research question* it was assumed that differences in speed would not only be found between curves, but also between different groups of drivers. With respect to the perceived characteristics, it was assumed that these match the potential differences found for behaviour. Differences in behaviour between participant groups were only found between different age groups. The fact that differences between age groups were found at all is astonishing because the sample used in this study was quite young. The differences were consistent throughout the different curve designs. In contrast to the assumptions, these differences in behaviour between age groups were not mirrored by differences in the perceived curve characteristics or by the expectations concerning appropriate behaviour for these curves. In combination with the behavioural differences between driver groups found between curves these results imply that demographic variables influence behaviour directly rather than indirectly via perception of the situation. The direct influence can be seen as a calibration factor that is applied to the ratings. It differs between driver groups and changes with changing resources.

4.3.5 Overall Conclusions of the Simulator Study With Respect to the Model

The driver and driving behaviour model for rural roads proposes that the perception of environmental characteristics and the expectations derived from this perception have a sequential influence on behaviour (Figure 14). Unlike the laboratory study, which analysed the broadest possible range of different rural road designs, the number of experimental variations in the simulator study was limited. Therefore, the influence of environmental variation was analysed indirectly by comparing differences in behaviour to differences in the subjective ratings. The assumptions were investigated in the simulator by

- experimentally changing the environment on straight road sections and by adding or removing elements which could serve as behavioural cues in curves;
- asking participants how they perceived the situation and what behaviour (here: the amount of speed reduction required) they regarded as appropriate; and by
- measuring characteristic speed parameters for the road sections under investigation.

The differences found for behaviour were then compared to differences in the ratings. It was found that

- the experimental variation of the environment on straight road sections did indeed influence speed, as did adding or removing behaviourally relevant cues at curves, and
- differences in the ratings for the perceived road and environmental characteristics partly mirrored the differences found for behaviour, which is interpreted as supporting the model assumptions.
- However, this was only the case for cues in curves, not for environmental variations on longer straight road sections.

Thus, the mechanisms accounting for the influence of environmental variation on behaviour are different for straight road sections and for curves. The reason for this difference is probably that speed must be adjusted in an open-loop way before entering a curve. To do so the driver must be aware of the characteristics of the oncoming curve. Because the anticipated estimation of the cuve geometry is difficult (Shinar, 1977) drivers must rely on additional cues.

In contrast to curves, environmental characteristics alongside straight road sections only play a role if they serve as cues for upcoming situations such as curves or crossroads. This was not the case for the variations that were examined on the straight road sections in this experiment.

If speed on straight road sections cannot be explained by differences in the perceived road characteristics, how else can it be explained? It was concluded in chapter 4.3.4.5 that this is achieved directly via the optic flow as proposed by Gibson (1986). In contrast to the findings for curves, this assumption conflicts with the model assumptions. For the model, this means that a direct path should be added from ‘objective road geometry and situation’ and ‘affordances and cues’ to ‘current behaviour (B_{C2})’ in the driver and driving behaviour model for rural roads.

With respect to the selection of marker items (see chapter 4.2.5.5), rated risk was indeed very successful in predicting behaviour. However, the same result was found for rated demand, supporting the idea that both variables are not distinguished by drivers when collected as ratings (see also chapter 2.3.7). ‘Monotony’ was not useful in curves, but might serve as a preliminary indication on straight road sections.

An additional aspect in the simulator study concerned the influence of demographic variables. Differences between drivers or driver groups are so far only indirectly integrated into the model via ‘knowledge, experience, and mental models’. It was assumed (see chapter 4.3.2) that differences in behaviour between driver groups are mirrored by differences in perception and expectations. If this assumption were supported, the integration of driver characteristics via ‘knowledge, experience, and mental models’ would suffice to explain behavioural differences between driver groups. However, the findings did not support this assumption. Behavioural differences were indeed found, namely between male and female drivers on straight road sections, and between age groups in curves. However, these behavioural differences were not matched by differences in the ratings.

The findings of the simulator study suggest that the model should be modified with respect to a direct path of perception to behaviour and with respect to the effect of demographic variables. This will be done at the very end of this thesis in order to allow the results of the field study to be integrated also. The field study deals exclusively with the effect of behaviour on accidents and therefore, the changes to the model outlined so far do not have direct consequences for the field study.

4.4 On-the-Road Driving Tests: Behaviour and Accidents

4.4.1 Introduction: Rationale Behind the Driving-Experiments

The on-the-road field study was conducted to analyse the final part of the model. This part deals with the effect of behaviour on accidents. If the assumptions formulated in the model can be confirmed, the propensity of road elements with respect to accident occurrence could be assessed before these accidents actually happen. With the help of additional simulator studies it might even be possible to do this in the planning stage of new roads which would result in a considerable benefit for road safety. In this respect, the approach used here goes beyond the approach used in the ESN (FGSV, 2003) or EURORAP II (Lynam et al., 2007) which uses existing accident data in partial combination with an assumed accident cost rate for the road class under investigation.

In the final part of the driver and driving behaviour model for rural roads, behaviour was defined as safe as long as it does not deviate from appropriate behaviour in the situation. The focus in the field study thus shifts from open-loop control, as in the laboratory and simulator study, to closed-loop control. In order to test the model assumption, ‘appropriate behaviour’ had to be defined. At present, the definition of appropriate behaviour in free-flowing traffic conditions is restricted to speed and lane-keeping (Reichart, 2001). For example, maximum possible speed in a curve can be defined with the formulas introduced in chapter 4.3.3.12. However, applying these formulas has some shortcomings. Firstly, not all parameters are usually known, thus requiring the use of assumed values which naturally affects the results. Secondly, such an upper limit does not take into account the fact that behaviour might be dangerous even if it is below this maximum possible value.

In the previous INVENT project[11], driving behaviour was classified into errors of different severity, depending on how close the behaviour was to this maximum value and how long it was maintained (Glaser et al., 2005; Nirschl et al., 2004). Because of the

---

[11] Further information at www.invent-online.de; (Sub-project: FVM / MMI; Work-package 3100: Development of an Assessment Procedure for IVIS and ADAS).

relationship between workload and performance (de Waard, 1996; Fuller, 2005), psychological variables were also included in this assessment (Weller, Schlag, & Nirschl, 2006). However, defining an appropriate level for these variables is even more difficult or even impossible (see the workload redline discussion, de Waard, 1996).

Therefore, for this study, an approach had to be chosen which could overcome these shortcomings. The solution was found by using a repeated-measures design, whereby curves with a high accident-rate were compared to geometrically similar curves with no accidents. The latter curve then served as a baseline condition for which the behaviour could be equalled to appropriate behaviour. This approach not only allowed the experimental testing of hypotheses, but also the exploratory comparison of parameters for which no hypotheses were defined in advance.

4.4.2 Hypotheses and Further Research Questions

The subsequent hypotheses were formulated for the comparison of two curves that were similar in geometry but different in their accident-rate.

Hypothesis 1:

The minimum speed in the high accident-rate curve is higher than in the low accident-rate curve.

This hypothesis is based on the fact that accident severity and accident probability increase with speed (L. Aarts & Schagen, 2006; Elvik & Vaa, 2004). Because maximum permissible speed is a function of geometry (see chapter 4.3.3.12) and both curves have more or less the same geometry, higher speeds can be interpreted as leading to both more accidents and more severe accidents. The reason for the higher speeds in high accident-rate curves can be seen in an open-loop underestimation of the curve difficulties or effects caused by the optic flow in the preceding straight road section. Both effects were shown in the preceding simulator study.

Hypothesis 2:

The maximum lateral acceleration in the high accident-rate curve is higher than in the low accident-rate curve, given that the driven radius is the same.

This hypothesis is based on the same considerations as Hypothesis 1: higher speeds lead to higher lateral acceleration and thus increase the risk of leaving the road due to the laws of physics (see chapter 4.3.3.12). Because the driven radius can deviate from the measured radius of the curve (Spacek, 2005), it had to be ensured that the same radii were driven. Because lane-keeping could not be measured with the measuring vehicle (see also chapter 4.4.3.7), a minimum experimental control of this factor was ensured by only comparing curves of the same curve direction.

Hypothesis 3:

Workload in the high accident-rate curve is higher than in the low accident-rate curve.

The task-capability interface model (TCI-model, Fuller, 2005) and other workload models (see chapter 2.3.5) propose that performance decreases and accident probability increases if workload surpasses a critical level. However, at present such a 'workload redline' cannot be defined exactly (de Waard, 1996). Therefore, higher workload must be used as the indication of a critical tendency towards the unknown workload redline. Higher workload in the high accident-rate curves can be explained by the higher speeds which were assumed for these curves (see hypothesis 1). These were in turn attributed to an underestimation of the anticipated curve difficulties before entering the curve.

Hypothesis 4:

The percentage change in workload between the minimum workload value before the curve and the maximum workload value in the curve is higher for the high accident-rate curve than for the low accident-rate curve.

In contrast to Hypothesis 3, which is based on the TCI model of Fuller (2005), Hypothesis 4 is based on workload homeostasis theory (Fuller, 2005; Gstalter & Fastenmeier, 1995). According to this theory, drivers regulate speed in order to keep workload more or less constant. A failure in workload homeostasis is regarded as an indication of a dangerous situation, similar to a failure in risk homeostasis (Wilde, 1994, 2001; Wilde, Claxton-Oldfield, & Platenius, 1985).

Maximum workload as used in Hypothesis 3 would be an inappropriate parameter to test workload homeostasis. This is because the target level of workload (here: the demand of the situation in combination with the speed driven) not only differs between individuals, but more importantly might also differ between situations, that is, it might change with time along a driven course. The percentage change between the minimum level of workload before the curve and the maximum level of workload in the same curve compensates such changes between different sections of a course.

It should also be noted that homeostatic control does not mean that the parameter under investigation is constant, but rather that it oscillates around a target value. Therefore, a value different from zero in the percentage change does not necessarily indicate a failure of homeostasis and will probably also be found in low accident-rate curves. Therefore, it is once again the comparison between high and low accident-rate curves together with the analysis of differences found between both which are required to test this hypothesis.

Hypothesis 5:

Maximum speed before the curve is higher for the high accident-rate curve than for the low accident-rate curve.

In this thesis, it was assumed that the reason for higher speeds in high accident-rate curves is underestimation of curve difficulties (see preceding hypotheses). Another explanation might be carry-over effects from the preceding road section. These could be due to the geometry of the preceding section, but also because of the optic-flow when geometry is the same. While, in general, care was taken that the geometry of the preceding road sections were the same within curve pairs, such statements are necessarily approximations. For example, parameter values such as curvature change rate depend very much on the elements which are included and thus change with the length of the road section chosen for the calculation of the parameter. Therefore, using measured speed is a better indicator than approximated speed calculated from varying geometric parameters (see the results for the simulator study).

Additional research question 1:

Are workload and speed related?

In addition to the hypotheses above, an additional research question deals with the relationship between workload and speed. This question is addressed in some of the hypotheses above related to accident occurrence and individual drivers. Here, the question is understood more generally and will be addressed with respect to the changing demand of several curves with different geometric characteristics. The research question is based on homeostatic theories as described in chapter 2.3.

These theories assume that speed is used to compensate for different demands in order to achieve a certain level of workload or risk. However, empirical support of homeostatic regulation with respect to road geometry is sparse (see chapter 2.3.9). One reason might be that the seemingly clear predictions of the theories are less easy to test empirically in driving. This is because speed itself can be seen as a primary measure of workload (de Waard, 1996; Gawron, 2008). Thus, strictly speaking, a variation in reaction time could also be used for a homeostatic control of speed. The only limitation to this interpretation is the physical limit of curve driving (see chapter 4.3.3.12).

In order to account for this potential trade-off, speed as the primary task measure and reaction time as the secondary task measure of workload have to be analysed jointly with respect to geometric demand. In order to align with the assumption of homeostatic theories, either speed or reaction time should be independent of geometric demand.

Additional research question 2:

Does the method used to assess workload in turn influence normal behaviour?

This research question is related to the preceding question. Depending on the method used to collect data on workload, the measurement procedure itself might be a source of demand and might therefore affect normal behaviour. This will almost certainly be the case if a secondary task is used to collect workload. However, strictly speaking, the dual task method requires that the primary task is not affected by the secondary task (Brown, 1978; Tsang & Vidulich, 2006). On the other hand, both tasks have to make use of the same resources (see the cube model of Wickens, 1984, 2008) because otherwise the secondary task cannot measure the spare capacity of the relevant resources which remain when the primary task is conducted. If and to what extent the secondary task influences primary task performance needs to be clarified.

The research paradigm used to analyse this question was to compare sections driven without PDT in one direction to sections driven with PDT in the other direction. A prerequisite for this kind of comparison is that the road sections which precede the experimental road section are comparable, regardless from which direction the driver approaches the experimental road section. Besides the question of the PDT-influence on behaviour, the results to this additional research question are important because they allow some insight into the dependent variables in general. Thus, the entire results contribute significantly to interpreting the results for the hypotheses.

Additional research question 3:

Does the gaze data show any characteristic differences between high and low accident-rate curves?

Although there is considerable evidence that longer fixation durations are associated with deeper processing and more demanding tasks (overview in Weller, Weise & Schlag, 2008, but see also Chapman & Underwood, 1998), it is difficult to formulate a hypothesis concerning fixation duration which takes into account the effect of the PDT on gaze behaviour. Such an effect is probably present, because reacting to the PDT requires the driver to look away from the road. Reacting faster and more often to the PDT diminishes the time available to the driver to look at the road. This reduces fixation duration, given that the same number of objects are looked at. However, if the number of objects which are fixated upon is reduced, for example, when only seemingly relevant information is looked at (Bartmann, 1990, cited in Weller, Schlag, Gatti et al., 2006), fixation durations could also be longer. Therefore, not only the fixation duration but also the distribution of gaze data across the scenery has to be analysed and compared. Several parameters were used here and are described in chapter 4.4.3.9. As with fixation duration, hypotheses for these parameters cannot be formulated when an additional detection task is present.

4.4.3 Methodology

4.4.3.1 General Course of Events

For this study, driving experiments were conducted with a vehicle equipped to measure driver and driving behaviour. The course was chosen so that it included road elements with a high number of accidents together with geometrically similar road elements with fewer or no accidents.

4.4.3.2 The Test-Route

The safety-related hypotheses required the presence of locations with a high number of driving accidents. Because of the nature of accidents (rare events, often happening ‘by chance’) road authorities define the number and type of accidents that must occur in a given time in order for a location to qualify as an accident black spot. When applying such criteria it must be taken into account that they differ throughout Europe (Elvik, 2008). In Germany (FGSV, 2001), accident conspicuous locations (German: ‘Unfallhäufungsstellen’) are defined as having either

- five or more accidents of the same accident type, recorded by the police but independent of accident severity within a 12-month period;
- five or more accidents of the same accident type with personal injury within a 36-month period; or
- three or more accidents of the same kind with severe personal injury within a 36-month period.

The time periods chosen allow phenomena such as a regression-to-the-mean to take effect but reduce the likelihood that accident figures are affected by uncontrolled influences such as a reconstruction of the road or changes in the AADT (see Elvik, 2002, for a discussion of such effects).

For the present study, 'driving accident' was selected as the most relevant accident type. Driving accidents are usually single-vehicle accidents such as run-off-the-road accidents and therefore can be attributed to the interaction of the driver with the road and road environment, independent of other drivers. Driving accidents are the most common accident type on rural roads and presently account for almost 40% of all accidents with personal injury on rural roads in Germany (Statistisches Bundesamt, 2006, pp. 59, 65, 249).

In former times, accident occurrence was visualised with pins of different colours stuck into maps. These maps were decentrally administered in every jurisdiction by the police and local road authorities, which made analyses difficult and the selection of a test route according to the above-mentioned criteria a rather time-consuming task. Today, accident data is administered centrally with computer programmes, although these programmes do in fact differ between the different federal states in Germany. In Saxony, the EUSKA (PTV, 2006) computer programme is used.

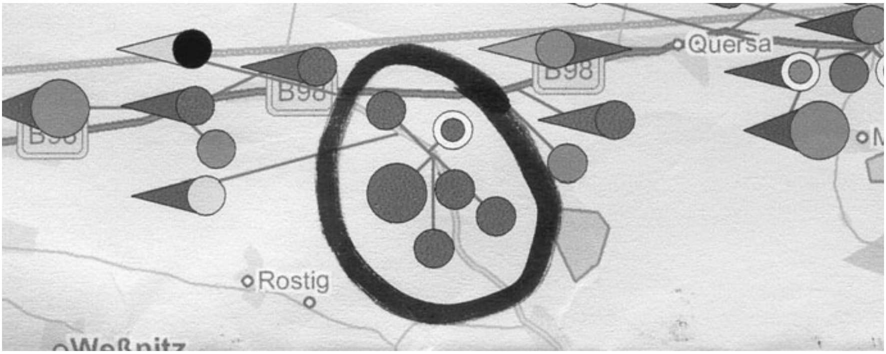
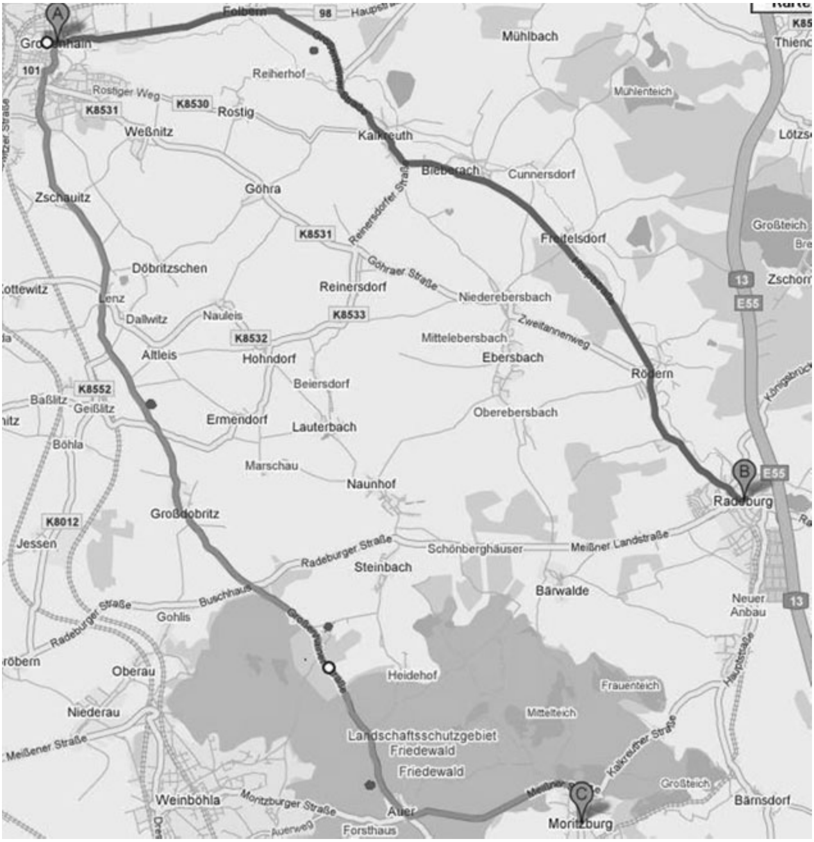


Figure 32. Print-out of a computer-based accident map (EUSKA, PTV, 2006) (encircled: section A09/A50).

A print-out of EUSKA for a location which was later part of the test route is shown in Figure 32. The accident location is fed into the database by the local police via a GUI-map. Although there is a designated variable in the database for the direction in which the car was driving at the time of the accident, this information is often missing or – according to several members of the Chair of Road Planning and

Road Design – unreliably coded in the database. This information could therefore not be used.



Source: Google Maps Deutschland.

Figure 33. Test-route used for the driving experiments (Radeburg, Großenhain, Moritzburg and back).

All programmes used in Germany to record accident data use the same symbols and the same terminology, which are defined in guidelines (for Germany, FGSV, 2001), and which are in accordance with the former accident maps. Every accident is visualised with a dot that differs in colour depending on accident type, and in size depending on accident severity. For example, the dots in the black marking in Figure

32 represent driving accidents of different severity. The largest one indicates an accident with severe personal injury, the smaller ones accidents with non-severe personal injury, and the smallest one with the white border property damage only.

For the selection of the test route, an area of approximately 50 km radius from Dresden was examined on the accident map to find locations with a high number of driving accidents. Because the computer-based administration of accident data has only recently been introduced in Saxony and Germany, the accident data which could be used for this area was restricted to the years from 2004 onwards. Furthermore, because the data for 2006 was not yet available at the time of selecting the test route, the analysis was restricted to the two-year period covering 2004 and 2005. Although it would have been preferable if a three-year period had been used, the accident map nevertheless indicated some accident black spots based on the FGSV-criteria named above.

For a preliminary inspection of their suitability, the RoadView computer software was used. This programme was developed and provided by the Chair of Road Planning and Road Design at TU Dresden (Dietze, 2007). Amongst other things, the database behind RoadView contained information on road geometry and pictures which were taken approximately every 10 m of the roads in the database.

After this preliminary inspection, some locations were discarded as not being relevant for this study. For example, this was the case for a curve with a very high number of driving accidents, which seemed suitable at first sight. However, this curve had an extremely small radius and was the first curve after an exit from an autobahn. Because of this combination of geometry and location it was firstly impossible to find a comparison curve, and secondly, this combination alone would have been enough to explain accident occurrence (Kämpfe, Schlag, & Weller, 2005; Kämpfe, Weller, & Schlag, 2004).

Other locations still regarded as relevant after this preliminary inspection or those not integrated in the database were inspected *in situ*. This step resulted in a further reduction in the number of available locations, mainly because they were undergoing reconstruction at the time of inspection. It is likely that the decision for this reconstruction was based on the same considerations which led to the selection of these locations for this study which is very positive from a road-safety perspective.

All in all, these steps considerably reduced the already small number of accident black spots which could be used for this study. Nevertheless, the route which was finally selected (Figure 33) still contained four such locations. This route was situated 20 km north of Dresden and consisted entirely of two-lane rural roads. The length of the course was approximately 40 km. As it was driven in both directions, the total distance amounted to 80 km. The four accident-conspicuous locations are situated at the four dots in Figure 33 and are described in more detail in the next chapter.

4.4.3.3 Characteristics of the Curves Used as Independent Variables

The main independent variable in the field study was the classification of locations as accident-conspicuous versus not conspicuous. This classification was based on the number of accidents as described in the preceding chapter. Pictures of all curves for both directions can be found in Appendix A 4.1, the curvature plan for two high accident-rate curves and the reference curve is shown in Figure 34, others are shown in the chapters where the results are presented.

Because the geometric curve characteristics are critical with respect to the research paradigm used, the parameters which were used are briefly introduced. In general, curves can be described with several parameters, of which radius is possibly the most common. In addition to radius, the curve length is an important variable. Both parameters can be combined to calculate the curvature change rate of a single curve (CCR or CCR_S, see Appendix A 1.4). This is especially important because curves are not only circular curves, but are usually a combination of transition curves and circular curve. CCR can also be calculated for entire road sections. However, because CCR changes greatly depending on which elements are included, sections for which CCR is calculated should be characterised by similar changes in the deflection angles (FGSV RAS-L, 1995). If this is ensured, CCR can be used to approximate V85 or other speed parameters (see chapter 4.3.3.13) (see also Ebersbach, 2006).

For the subsequent enumeration, two CCR values in gon/km are given: the CCR_S-value that only includes the parameters of the singular curve and transition curves; and the CCR for the entire section for which some of the dependent variables were calculated (see next chapter). Because each location was driven twice, once in the outbound direction and once in the inbound direction, the same location was given different names. The names are a combination of the prefix 'A' and a number. The numbers ascend with driven distance; numbers below 29 denoting the outbound direction and numbers from 29 onwards denoting the inbound direction. The numbers were not only given to elements reported in this thesis, but also to elements which might be used for future analyses. The subsequent enumeration gives the characteristics of the four high accident-rate locations. A table of the accident details is provided in Appendix A 4.2.

- Section A09/A50 (see also Figure 32) was characterised by one severe personal injury accident, four personal injury accidents, and one property damage only accident. All accidents happened when the road surface was wet, two out of the six at dusk, the rest in daylight. The geometry of the single curve was: radius = 128 m, $CCR_S = 359.0$, $CCR_{A09} = 157.1$, $CCR_{A50} = 144.8$.

- Section A14/A44 was characterised by four severe personal injury accidents and one property damage only accident. Three out of the five accidents happened on a wet or icy surface, four out of the five in daylight conditions. The geometry of the single curve was: radius = 129 m, CCR_S = 276.9, CCR_{A14} = 118.8, CCR_{A44} = 120.2.
- Section A24/A33 was characterised by three severe personal injury accidents and three property damage only accidents. Five out of these six accidents happened on a wet surface, one out of the six in darkness. The geometry of the single curve was: radius = 141 m, CCR_S = 451.5, CCR_{A24} = 120.4, CCR_{A33} = 239.3 (CCR_{A33} was very high because only a short straight road section was included due to a village).
- Section A20/A37 was characterised by two driving accidents with severe personal injury and one driving accident with non-severe personal injury. All accidents happened on a wet surface and during daylight. The geometry of the single curve was: radius = 198 m, CCR_S = 178.0, CCR_{A20} = 77.8, CCR_{A37} = 77.3.

Throughout the entire road sections in which the high accident-rate locations and the reference curve (see below) were situated, speed was restricted to 70 km/h in wet surface conditions only. An exception was location A09/A50, for which speed was limited to 70 km/h regardless of the weather condition, and also exclusively for the curve itself, not for the entire road section in which this curve was situated. Implications of the accident characteristics and the speed limits for the study are discussed in the next chapter.

For Hypotheses 1 to 5, at least one direction of these high accident-rate locations was compared to a single low accident-rate location with no accidents at all, namely location A21/A36 (see Figure 34). Before details of matching high accident-rate curve to reference curve are given, the pairs are shown in the following overview:

- section A14 was compared to the reference section A21;
- section A24 was compared to the reference section A21;
- section A37 was compared to the reference section A21;
- section A33 was compared to the reference section A36; and
- section A50 was compared to the reference section A36.

The reference section A21/A36 contained a single curve with a radius of 135 m and a CCR_S value of 471.6 gon/km. The CCR value of the entire section was comparable for both directions ($CCR_{A21/A31} \approx 80$ gon/km) due to a preceding curve with a large radius in each direction. When driven in the outbound direction, the singular curve with R = 135 m was a left curve (then termed section A21), and a right curve when driven in the inbound direction (termed section A36). The village after section A21 could not be seen until the curve was driven to its end (see also Appendix A 4.1). Although section A36 was situated after a village (see Figure 34), it could still be used as the reference curve for two other sections.

One of these sections was section A33, which was also located after a village. Unlike section A36, a speed camera was positioned in the village before section A33. This speed camera was located at a distance of approximately 400 m before the beginning of the curve in section A33 (see Figure 34).

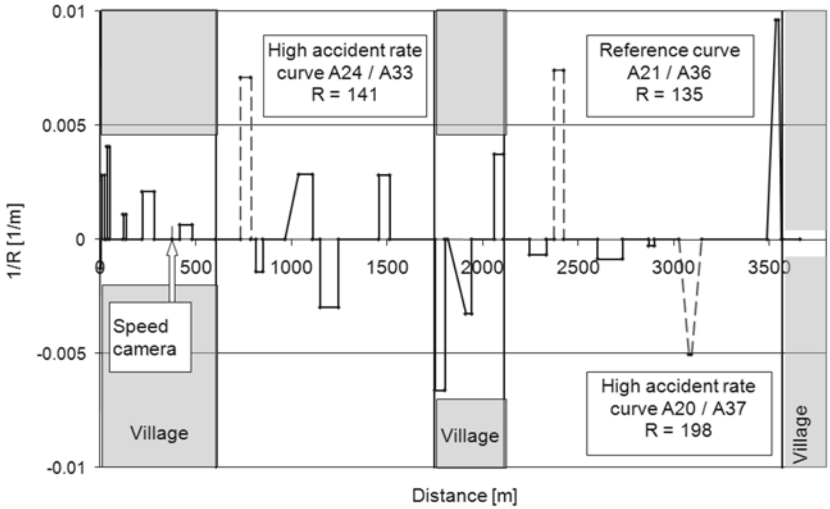


Figure 34. Curvature plan of sections A20, A21, and A24 when driven in the outbound direction from right to left, and of sections A33, A36, and A37 when driven in the inbound direction.

According to the responsible authority, this camera was already present before the period for which the accidents were analysed. Furthermore, it was positioned in the village to increase pedestrian safety (there were no pavements in the village), not because of the accidents after the village. Nevertheless, its presence likely had an influence on speed, even up to the beginning of section A33 (for the effects of speed cameras, see Elvik & Vaa, 2004; Lipphard, 2005), and should be taken into account when discussing the results. On the other hand, including both directions of location A24/A33 in the analysis is regarded as important because of the missing information concerning the direction in which a driver was headed at the time the accident happened (see chapter 4.4.3.2). By including both directions, this missing information no longer plays a role.

The second high accident-rate curve for which the curve in section A36 was used as the reference curve was the curve in section A50. Originally, only section A09, which corresponded to the outbound direction of curve A50, was planned to be used as the experimental curve. However, in the end, section A50 had to be used because section A09 was excluded from the analyses due to insufficient training with the PDT (see chapter 4.4.3.11). Section A36 was selected as the reference curve because no other reference curve was available with a corresponding radius, speed limit and curve direction to the curve in section A50. The idea behind using A36 as the reference curve was that the village before section A36 and the speed limit in section A50 might have had similar effects on the speed before the curve. Speed in the curves themselves could thus differ according to the assumed differences. Of course, analysis of differences between sections A50 and A36 can only be performed exploratory and its results can only be interpreted in the context of the other comparisons.

In general, comparing curves with radii ranging from 128 m to 198 m to a single reference curve with a radius of R = 135 m (see above) is far from perfect but could not be avoided because more suitable curves were not available within the driven course. The limitations described so far have to be taken into account when interpreting the data. All sections described so far were driven with a peripheral detection task (see chapter 4.4.3.8 below).

For additional research question 1 ('Are workload and speed related?'), all sections throughout the entire course were used which were driven with PDT; details are given in chapter 4.4.4.7. Further analyses were conducted with relation to additional research question 2 ('Does the PDT influence behaviour?'). These analyses were conducted for two sections that were compared to themselves when driven from the opposite direction. These two sections were driven with the peripheral detection task (PDT) in one direction and without PDT in the other direction:

- Section A14 was driven with PDT in the outbound direction and was compared to itself when driven without PDT in the inbound direction (= section A44).
- Section A20 was driven without PDT in the outbound direction and was compared to itself when driven with PDT in the other direction (= section A37).

Both curves were high accident-rate curves; further details of which were already described above. The analyses of differences when driven with versus without PDT might therefore also be used in relation to the hypotheses.

4.4.3.4 Implications of Course and Accident Characteristics for the Study

In the preceding chapter it was stated that speed for wet surface conditions was limited to 70 km/h in the experimental and reference sections. This limitation is likely due to the characteristics of the accidents which happened in these sections: most accidents happened on wet or icy surfaces (17 out of 20, equalling 85%) and during daylight (16 out of 20, equalling 80%, for section-wise details see the preceding chapter). The share of daylight accidents in the sample is somewhat larger than the corresponding share in the population. In Germany, 67% of all personal injury accidents together with major property damage accidents happened during daylight on roads outside urban areas without motorways (Statistisches Bundesamt, 2007, p. 95). While the differences for the factor 'light condition' can probably be explained by natural variation, the differences for the factor 'surface condition' certainly cannot. With the specification of accidents given above for the factor 'light condition', 62% of accidents happened on dry roads in the population, whereas it was only 15% in the sample.

It is obvious from the preceding paragraph that the combination of road surface and weather condition is likely to be specific for the accidents within the test-route. In order to take into account these specific characteristics, the experimental drives could have been conducted exclusively during wet surface or rainy weather conditions. However, this was not possible because it would have required an unusual flexibility concerning availability of test-drivers and the equipped vehicle. Furthermore, there were very few rainy days because the experiments were conducted during summer. In view of this, another approach was chosen. The experimental drives were conducted exclusively in dry weather conditions, thus keeping weather and surface condition constant.

Why was this approach useful? First of all, the speed limit for wet weather conditions was the same for all experimental sections including the reference curve without accidents. This means that differences in the accident figures between the curves must be attributed to the curve characteristics. Because of the close relationship between speed and run-off-the-road accidents in curves, this also applies to differences in speed. Thus, despite the formal speed limit, differences in accidents and speed cannot be explained by a general propensity of drivers to comply or not comply with the speed limit. Using the denomination of an analysis of variance, it can only be explained by a main effect of the curve characteristics, or an interaction effect between compliance with the speed limit and curve characteristics.

However, a main effect of curve on speed should also be found in dry weather conditions. An exception to this assumption would be the presence of an additional interaction effect between weather condition and curve with drivers driving slower in one curve than in the other curve, but only in wet weather conditions. This is as unlikely as it is that drivers comply with the speed-limit in wet weather conditions for one curve, but do not comply with it in another curve. The reason is that the driver has to infer the appropriate speed from the perceived curve characteristics. The only feature of a curve that changes in wet weather conditions is the road surface and an associated change of the coefficient of friction which, amongst other things, determines the maximum possible speed (see Appendix A 1.3). However, because the road surface was the same throughout the entire course, there was no reason for the driver to assume any differences in the coefficient of friction between the curves, regardless of weather condition. Any other differences in the perceived curve characteristics should thus have influenced driving behaviour to a similar extent, independent of the weather conditions, thus supporting the decision to conduct the experiments in dry weather conditions.

4.4.3.5 Preparation of the Raw Data for Statistical Analysis

The programmes which were named in chapter 4.1 were also used for the preparation of the raw data for the statistical analyses described in the next paragraphs.

Firstly, data were aggregated for sub-sections with a length of 25 metres within each experimental section. The parameter which was used for this aggregation step depended on the variables (see also next chapter). For speed, fixation duration, and reaction time, the average was used, whereas the maximum was used for acceleration. Further variables of gaze such as the number of fixations (see chapter 4.4.3.9) were exclusively calculated for these sub-sections. This approach was chosen because it allowed a detailed analysis of exactly when and where differences between two or more curves become evident, while at the same time the amount of data could be considerably diminished. This approach proved useful in previous studies (Lippold & Schulz, 2007) and was therefore also integrated into the programme used to analyse the gaze data (see chapter 4.4.3.9).

An additional processing step was required for the reaction-time values when used for the graphical representation of the data derived from the first aggregation step. This was necessary because reaction times longer than the time required to drive a sub-section could result in ‘missing’ values for the preceding sub-section. These values are not real missing data but rather represent the ongoing reaction process. Nevertheless, the values were interpolated for each participant in order to show correct reaction-time values for each sub-section. Because this interpolation is only relevant for the graphical representation of the data as shown in Figure 39 and Figure 40, but not for the statistical tests, the algorithm used for the interpolation is described in Appendix A 4.3 rather than here.

The second step before statistically analysing the data was performed by further aggregating the values of the first aggregation step. The parameters which were used for this second aggregation step were minimum, maximum, average, and percentage change between maximum and minimum value (see next chapter). The second aggregation step was conducted in order to overcome certain shortcomings in the analysis of the data by sub-sections. While such sub-section-wise analysis is useful to describe behaviour in single curves, it is less appropriate when different curves are to be compared. This is due to the fact that in this case, the results will depend greatly on which sub-sections are matched. Of course, such matching would be performed via a reference point of each curve as described in Weller & Schlag (2007).

However, defining a reference point is not as straightforward as might be expected when curves of different geometric characteristics are to be compared. If the curve apex is used as the reference point, the presence or absence of a clothoid in one of the curves influences the results. The same applies when defining the curve's beginning, whereby a limit must be chosen below which clothoids are regarded as part of a curve or not. Analysing the data by sub-sections, yields a large number of results which are highly influenced by how the reference point is defined and which reference point is chosen. Using the second aggregation step allowed curves or their preceding road section to be described with a single characteristic value for each participant and thus to overcome the shortcomings of the preliminary analysis.

4.4.3.6 The Dependant Variables: Overview

Based on the rationale of the study, the most important dependant variables were speed and workload. Speed was measured with the experimental vehicle (see next chapter), which also logged lateral and longitudinal acceleration. The method chosen to assess workload was a peripheral detection task (see chapter 4.4.3.8). For both speed and reaction time, the minimum, the maximum, and the average values were calculated as the result of the second aggregation step described in the previous chapter.

Because the experimental sections were usually a combination of a more or less straight road section before a singular curve and the singular curve itself (see chapter 4.4.3.3), maximum speed and minimum reaction times were typically found on the straight road section, whereas minimum speed and maximum reaction time correspond to the curve itself. The parameters were assessed within a section covering approximately 150 m before the curve beginning up to approximately 50 m after curve ending. In addition, the percentage change between the maximum and the minimum value was calculated as:

$$\text{Percentage change} = \frac{\text{maximum} - \text{minimum}}{\text{maximum}} \times 100 \tag{20}$$

Strictly speaking this is not correct, because speed decreases from the maximum to the minimum whereas reaction time increases from the minimum to the maximum. However, in order to ensure a certain level of comparability for the height of the parameter between both variables in order to assist with interpretation (see for example chapter 4.4.4.7), the calculation from Formula (20) was preferred. Concerning lateral acceleration and longitudinal deceleration, only the maximum values were used in this thesis because they alone are relevant to accident causation. The variables and parameters are also summarised at the beginning of the chapters where the results are discussed, together with peculiarities of singular cases or further details wherever necessary.

In addition to the variables named above, gaze data was assessed with a contact-free eye-tracker which was integrated into the measurement vehicle (see next chapter). Details concerning variables of gaze data and the parameters used are provided separately in chapter 4.4.3.9.

For this study, lane-keeping was not used as a dependent variable because it could not be measured precisely enough with the measurement vehicle at the time of data collection. However, because of the prominent role of speed in accident causation on rural roads (Lippold, Dietze, Krüger, Scheuchenpflug, & Mark, 2005; Statistisches Bundesamt, 2007) this is not seen as problematic.

4.4.3.7 The Measurement Vehicle

The car used to assess the dependant variables was the measurement vehicle of the Chair of Road Planning and Road Design at TU Dresden (see Figure 35). Besides parameters of driving behaviour such as speed, and lateral and longitudinal acceleration, driver behaviour was assessed with a peripheral detection task. Gaze behaviour was recorded with the integrated, contact-free eye-tracking system 'Smart Eye' (two-camera system, see upper right picture of Figure 35). The scenery camera integrated in the eye-tracking system was used to record each drive on video (middle right picture of Figure 35).

All data were recorded with time stamps synchronised and referenced to the location which was available through the high-precision GPS-based system APPLANIX. More detailed descriptions of the measurement vehicle can be found in Lippold & Schulz (2007).

Figure 35. Experimental vehicle of the Chair of Road Planning and Road Design at TUD (Prof. Lippold).

4.4.3.8 The Peripheral Detection Task

A peripheral detection task (PDT) was chosen for the measurement of workload in the field study. The PDT was chosen for several reasons. Firstly, it is an objective method which overcomes several shortcomings associated with using subjective data collected with rating scales (Bortz & Döring, 2006 for an overview). Secondly, its more or less continuous presentation (see below) allows a workload profile of a road section to be determined. In contrast to physiological data, reactions to changes in demand are directly measurable and are not subject to delays caused by the physiological system (Schandry, 1998 for an overview). Also in contrast to physiological data, the PDT can

be regarded as a robust method that minimises the occurrence of missing data due to measurement errors or unsystematic errors due to the environment.
The PDT chosen for this study required a reaction to signals shown on a 19 inch touch screen monitor. The screen was installed near the dash board to the right of the driver. The position of this screen was similar to the position of the laptop shown in the lower right picture of Figure 35. The screen was adjusted individually for each participant via a number of flexible joints in order to be easily reached. The signals presented to the participants were red rectangles (55 × 40 mm) which popped up at different locations across the whole screen. The location of the rectangles was quasi-random as they were taken from a file of previously randomly created positions. The reaction required from the participants was to touch these rectangles, which caused the rectangles to disappear. The time between onset of the rectangles and reaction was measured as reaction time (RT) and served as a proxy variable for workload.

The time between reaction to one rectangle and onset of the next rectangle was parameterised. This was done to prevent individual reaction strategies from influencing reaction-time results. The time chosen for this interval had to serve three purposes: firstly, it had to ensure that drivers could drive safely with enough time to take up visual information needed for safe driving. Secondly, the drivers should feel sufficiently stressed by the task itself to result in meaningful reaction-time values. Finally, the statistical analysis of fixations directed at the scenery should still be possible even when the PDT was being attended to. A decision concerning the duration of this interval was based on an analysis of the literature.

Schweigert (2003) found a modal value for fixation durations of between 300 to 400 ms in driving studies in the field. Mean fixation durations were between 450 ms without specific objects present and 730 ms when directed at a car ahead. In simulator studies, Velichkovsky et al. (2002) distinguished between preattentive fixations and attentive fixations. Preattentive fixations were defined as being at or below the modal value of fixation durations which they found to be 204 ms. Attentive fixations were defined as being at or above the mean fixation duration, which was found to be 400 ms.

Based on these results, the interval chosen was 500 ms. This value allowed drivers to detect hazards if they were present, but prevented them from fixating on irrelevant objects for a longer time. Before the final application in the study, this interval was successfully tested in preliminary test drives on real roads.

The PDT had to be performed by the participants at selected locations. Start and stop instructions for entire sections were given to the participants by the instructor, who was seated in the back seat of the car. The participants performed the task at a standstill before driving and during the first 500 m of the test-drive. After 10 km of driving with several additional PDT sessions, the participants reached the first section relevant for this study (section A09).

4.4.3.9 The Gaze Data

Gaze behaviour was recorded with the contact-free two-camera solution of Smart Eye, which was already integrated into the measuring vehicle at the time of the experimental drives (see chapter 4.4.3.7). Contact-free eye tracking has three major advantages:

- the driver is not constantly reminded that eye-movements are being tracked;
- longer drives are possible because an uncomfortable helmet with cameras does not have to be worn by the driver; and
- there are no liability issues because of additional danger to the eyes in the case of an accident.

The system used the cornea-reflex method, which is essentially based on the fact that light directed towards the eye is reflected from the cornea in a direction which depends on the position of the cornea. In order to compensate for different light conditions, the system was equipped with infrared light emitting diodes which were directed towards the eyes and were placed at the eye-tracking cameras. The system had to be calibrated for each participant in order for the system to identify the eyes. In a second step, the scenery camera was calibrated to the gaze data by asking participants to fixate pre-defined objects in the scenery. Both calibration steps were conducted with each participant prior to the experimental drive at a car park. Depending on the participant's facial features, which define suitability for eye tracking, both calibration steps took approximately 30 minutes. Data were recorded with a frequency of 30 Hz.

In the original Smart Eye log-file, each gaze data point was represented by the X-, Y-, and Z-components of a normalised vector. This vector in a three-dimensional space had to be projected onto a two-dimensional plane in order to be visualised in the video. This was performed in the background by Smart Eye. The location at which the vector penetrates the plane was depicted as a cross and was superimposed on the video of the scenery camera. This video was presented on an additional monitor in the back of the car and recorded on VCR. On the monitor, the experiment leader could check the quality of eye tracking in real time during the experimental drives. Furthermore, the recorded video with gaze location allowed object-based analyses of gaze data, if required.

The gaze position on the video was not directly recorded in the log-file, but had to be calculated anew with the algorithms used by Smart Eye and the information stored in the calibration file of the scenery camera. This is due to the fact that the combination of the eye-tracking cameras and the scenery camera was a custom-made solution that required several technical issues to be solved by the Chair of Road

Planning and Road Design in cooperation with Smart Eye. Because these cannot be discussed in this thesis, interested readers are referred to Lippold & Schulz (2007) and a student paper by Weichert (2005) where they are discussed and explained in detail.

The gaze data were analysed using the software WatchOut (Schulz, 2007), which was tailor-made at the Chair of Road Planning and Road Design at TU Dresden for the hardware in the measurement vehicle. Because the algorithms used by this software proved successful in a former study (Lippold & Schulz, 2007), it was decided to use WatchOut instead of programming own software tools. WatchOut uses several variables and parameters of gaze behaviour. For this thesis, the subsequent variables and parameters were used for the analyses and are described in detail below:

- fixations (number and duration);
- the scan path;
- the standard deviation of the gaze location; and
- the spatial density index.

All parameters were calculated and averaged by WatchOut for sub-sections of 25 metres within the experimental sections (see chapter 4.4.3.6).

The definition of fixations requires several decisions concerning temporal and spatial thresholds of the gaze data (Goldberg & Kotval, 1999; Salvucci & Goldberg, 2000). The value of each threshold can in turn influence the outcome (Karsh & Breitenbach 1983, cited in Lippold & Schulz, 2007; Salvucci & Goldberg, 2000). In the aforementioned student paper (Weichert, 2005), fixations were preliminarily defined by using an algorithm developed by Jacob (1995, p. 273):

> [This algorithm] (…) watches the input data for a sequence of 100 milliseconds during which the standard deviation of the reported eye position remains within approximately 0.5°. As soon as the 100 ms have passed, it reports the start of a fixation and takes the mean of the 100 ms worth of data as the location of that fixation. (…) Further eye positions within approximately one degree are assumed to represent continuations of the same fixation (rather than a saccade to a new one). To terminate a fixation, 50 ms of data lying outside one degree of the current fixation must be received. Blinks or artefacts of up to 200 ms may occur during a fixation without terminating it.

This definition was adjusted to the measurement system in real driving conditions and resulted in the final definition which was applied by WatchOut (see also Lippold & Schulz, 2007, Weichert, 2005):

- The start of a fixation is recorded when gaze direction lies within an area defined by 0.6 degrees for a duration of 100 ms (three data points with the 30 Hz used by Smart Eye).
- The averaged gaze data location is defined as the fixation location.

- The end of a fixation is defined as when gaze direction is outside 1.6 degrees of this fixation location for more than 100 ms.

With the assumed duration of one second needed to drive a sub-section and an average fixation duration in driving of approximately 400 ms (Velichkovsky et al., 2002, see above), approximately two to three fixations per sub-section could be expected for the data at hand. If fixations started at the end of one sub-section but were terminated in the next sub-section, the fixation was assigned to the former only, with the duration of this fixation also including the time in the latter sub-section. Thus, data were not lost despite the division of the sections into sub-sections.

Based on the fixation locations, the scan path was calculated as the Euclidian distance between two consecutive fixations and was averaged for the sub-sections. Thus, the scan path took into account the temporal succession of fixations.

In order to be independent of the thresholds used to define fixations, additional parameters were used which were not based on fixations but on the raw data. One parameter was the standard deviation of the gaze data. In a first step, the centre of the gaze data was determined by averaging the X-values and by averaging the Y-values:

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}; \overline{Y} = \frac{\sum_{i=1}^{n} y_i}{n} \tag{21}$$

where n equals the number of data points.

In a second step, the distance to this centre was calculated for each data point as:

$$d = \sqrt{(\overline{X} - X_i)^2 + (\overline{Y} - Y_i)^2} \tag{22}$$

this distance was averaged as:

$$\overline{d} = \frac{\sum_{i=1}^{n} d_i}{n} \tag{23}$$

and the standard deviation was calculated as:

$$SDXY = \sqrt{\frac{\sum_{i=1}^{n}(d_i - \overline{d})^2}{n-1}} \tag{24}$$

The interpretation of this parameter is similar to the interpretation of the standard deviation as described in Victor, Harbluk & Engström (2005) where the authors used glances as defined in (DIN EN) ISO 15007-1 (2003) instead of the raw data, and degrees instead of pixels as the unit of the standard deviation. While glances are appropriate to assess the effect of in-vehicle devices, as defined, for example, by the European Commission (2007), the use of raw gaze data is more appropriate for driving on rural roads. This is because driving is to a large degree governed by pursuit movements and peripheral visual information uptake which takes place independently of whether a fixation or glance is present or not.

Finally, a spatial density index was calculated. The spatial density is one of four parameters which are usually used to describe the spatial distribution of gaze data (Rötting, 2001). Recently, a new parameter has been developed in addition to these four (Weiße, 2005; Weller, Weiße, & Schlag, 2008). This latter parameter combines the average and the variability of the spatial distribution within a single index, mainly in order to overcome difficulties which arise if more than a single centre of fixation is present. All these indices have in common that the area in the X- and Y-direction within which fixations are recorded is at first subdivided into smaller fields. The spatial density index is then defined as the share of those areas at which fixations are directed (Goldberg & Kotval, 1999). Neither the number of fixations on a field nor the duration of these fixations is taken into account.

The fields defined by WatchOut had a height and width of three degrees from a distance which resulted in 20 fields in the X-direction and 16 fields in the Y-direction. The resulting 320 fields covered the area surveyed by the scenery camera. This area was smaller than the area surveyed by the cameras which recorded the eye movements but still covered the road ahead and the scenery to the left and right of the road (see Figure 37). In contrast to the method described in Goldberg & Kotval (1999), WatchOut did not use fixations to count fields, but the raw data as used for the calculation of the preceding parameters. In general, an increase in the value of the spatial density index indicates that more fields have been ‘looked at’. Thus, the name spatial density index is somewhat misleading. Typical averaged values were situated around 0.03, which means that an average of 9.6 fields were looked at within this specific sub-section (9.6 / 320 = 0.03). Whether a higher spatial density index (i.e. more fields which are looked at) also means that gaze data were spread more widely across the scenery can only be ascertained when the standard deviation of the data and the scan path are analysed and interpreted jointly with the spatial density index.

Preliminary analysis of the data for the spatial density index revealed that, although raw data were used instead of fixations, sub-sections existed for which no spatial density could be calculated, that is, the spatial density index was zero. This was due to the PDT, which resulted in a higher proportion of gaze data outside the area covered by the scenery camera. This area was smaller than the area covered by the eye-tracking cameras (see above). Therefore, in order to compare entire sections, the parameters of the spatial density index used were the maximum or the average, instead of the minimum value. Of course, the restriction of the spatial density index to the area covered by the scenery camera diminished potential differences between sections driven with versus without PDT. However, these differences should still be visible in the data.

For the other variables of the gaze data, the parameters maximum and average were used for the statistical analyses (see also chapter 4.4.3.5). The minimum values were not used because aggregating the values for the minimum resulted in a value of zero for most participants in most sections which made a comparison between two or more curves based on the minimum inappropriate. The zero values were present in almost all curves somewhere in one of the sub-sections whenever drivers looked away from the cameras for a time period longer than it took to drive the sub-section. With a speed of 90 km/h, which equals 25 m/s, it takes exactly 1 s to drive a sub-section of 25 m. Such turning away of the participants from the cameras was mainly caused by the PDT. Because gaze behaviour only played a secondary role in this study (see chapter 4.4.2), discarding the minimum was accepted.

4.4.3.10 The Sample

The driver-sample consisted of 16 participants aged between 25 and 47 ($M = 30.6$, $SD = 5.3$) who were all employees of the TUD. The participants were recruited via mailing-lists of the TUD and via personal contacts. All participants had a valid driving licence. The average annual mileage driven over the last three years was 15,000 km/year (min. = 5,000; max. = 40,000; SD = 10,000). Furthermore, only participants who did not wear glasses were selected. This is because in preliminary studies the eye tracker was found to deliver data of higher quality for participants without glasses (Lippold & Schulz, 2007; Weichert, 2005, see also preceding chapter). For insurance reasons, the participants had to be recruited from the workforce of the TUD. The studies were conducted during daylight and normal office hours with as little traffic as possible. This required the participants to leave their current work for the time of the experiments, which usually added up to four hours when the time needed to drive to and from the experimental course was included. The restrictions mentioned above naturally resulted in considerable difficulties in recruiting participants. These difficulties were acerbated by the fact that the experimental vehicle was only available for a short time period because it was needed for other studies before and after the field study.

4.4.3.11 Ensuring Data Quality

In empirical research, two criteria are used to describe the validity of a study (Bortz & Döring, 2006, see also chapter 4.3.3.8):

- internal, and
- external validity.

With driving experiments conducted in the field, external validity is less an issue of the situation and more an issue of sample characteristics. These were described in the previous chapter. It should be noted that the sample was not a representative sample of the driver population in Germany, and even less of Europe. Because the sample consisted of drivers who would be classified as low-risk drivers (middle-aged academics, no young male beginners or drivers older than 75), potential differences between high and low accident-rate curves might even be more pronounced in a representative sample. Therefore, significant results found for the sample used here will almost certainly be found for a representative sample. Conversely, non-significant differences might not only be a result of the small sample size but also of the sample characteristics. Therefore, the issue of external validity is not regarded as a major problem for this study.

In contrast to external validity, insufficient internal validity is usually an issue in field studies. This is because, due to manifold external influences, it is difficult to ascertain that the independent variable is the sole or relevant cause for changes in the dependant variable. For this study, three major sources of unwanted influences were identified:

- variation due to other traffic participants and temporary situational influences;
- insufficient practice with the peripheral detection task (PDT); and
- issues relating to the eye-tracking data.

External influences

The issue of temporary situational influences was controlled by conducting the experiments during similar weather and light conditions (see chapter 4.4.3.10). The issue of other traffic participants required more effort. In the case of slower traffic ahead, the vehicle ahead was either overtaken, or the experimental vehicle slowed down in the next village until the slower vehicle ahead had gone far enough ahead to allow the next section to be driven in free conditions. When the slower vehicle was overtaken, the data was not used from this section. Data was also deleted from the database whenever such control was not possible during the experiments. This was done because car-following constitutes a task of its own and fundamentally changes the driving task (Crundall, Shenton, & Underwood, 2004a, 2004b).

The deletion of the affected data was done section-wise after the experiments by analysing the videos. Data was deleted if the following were present:

- other traffic in front of the ego-vehicle, either in close proximity or when approaching (= decreasing TTC), or
- cyclists, pedestrians or construction works on the road.

Although the selection was based on the videos only, the situations allowed an unambiguous decision as to whether the situation could be kept or not. The sections deleted from the database are given participant-wise in Appendix A 4.4. It should be noted that each section deleted also resulted in the exclusion of the comparison section for this participant. This is because of the statistical methods for paired samples or repeated measures used to analyse the data. While SPSS offers the possibility of replacing missing values with the average value, it was decided not to use this function. The reason for this is that the effect on the results of using replaced missing values is far from clear, especially for variables showing a high inter-individual variation in combination with the small sample size used for this study.

Issues relating to the PDT

In a second step, the reaction-time values were analysed. Two issues had to be addressed:

- issues in relation to the specific PDT used here, and
- the question of insufficient practice with the PDT.

The first issue was related to how the PDT was programmed and how participants dealt with the PDT. In order to minimise susceptibility to technical error, the PDT-programme was started only once at the beginning of the drive instead of before each PDT-section. Start and stop instructions for the participants were given section-wise by the experiment leader. Even between two PDT-sections, the PDT programme calculated the time between onset of the last signal in the former PDT-section and reaction to this signal in the next PDT-section. These reaction-time values which reflect the time driven without PDT and indicate the start of a new PDT-section were deleted first.

Of the remaining values, those above the average plus three times the standard deviation, both across all participants and all sections, were further analysed. This analysis was conducted with regard to the location and by analysing the video. It turned out that some of these high values were situated right after the beginning or at the end of the PDT-section. Because these locations were usually straight road sections, the high reaction-time values did not reflect road geometry but can be attributed to the participants having to adjust to starting the PDT anew. All reaction-time values that could be attributed to these starting and ending costs were deleted. However, because participants were instructed to start the PDT before and end it after the section which was later analysed, these values were usually not needed for the statistical analyses.

The analysis of the videos of the remaining unusually high reaction-time values did not reveal an external cause and were kept. However, some of these values seemed to be associated with oncoming traffic at the time the curve was negotiated. While this increases the danger in this curve and might therefore also influence the PDT results, these values were kept. This was done because oncoming traffic was sometimes also present for other participants and this did not result in an unusual increase in reaction time. Furthermore, no clear definition could be found as to when oncoming traffic did or did not influence reaction-time values, neither with respect to the characteristics of the oncoming traffic (how far away; cyclists, car or trucks; how much), nor with respect to reaction time.

Deleting all sections with oncoming traffic, regardless of reaction-time values, would of course have been the best solution but could not be done because of the already small sample size. The reaction-time values of the subsequent sections might have been affected for some participants due to oncoming traffic: the values for one participant in sections 50 and 56, for two participants in sections 09, 14 and 21, and for three participants in section 24. However, as has been stated, no systematic influence could be identified with respect to oncoming traffic and therefore traffic condition is interpreted as generally increasing variation in the data, but not as systematically influencing the results.

The second issue in relation to the PDT concerned the question of sufficient practice. This question is similar to the question of familiarization with the driving task in the simulator study and could have influenced data quality to a similar extent. Following the rationale of the dual task method, the reaction-time values should reach a baseline within a training session from which point on reaction time (RT) only increases with additional task demand. Therefore, the participants in the driving experiments practised the PDT during standstill of the car before the experiments began, and – more importantly – during driving in several sections before approaching the first experimental section A09. Due to the simplicity of the task, this training was assumed to have been sufficient.

Could this assumption be confirmed by the data? Answering this question proved to be more difficult than would have been the case if the task had been conducted in a laboratory. First of all, the PDT was interrupted when driving through villages in order not to endanger pedestrians. Secondly, effects of insufficient training are hard to distinguish from effects of actual demand because demand changes with the situation. Thirdly, performance in the PDT is not only a question of capability but even more a question of setting priorities. These preferences do not have to be stable within a participant but might change with experience, situation or state of the participants.

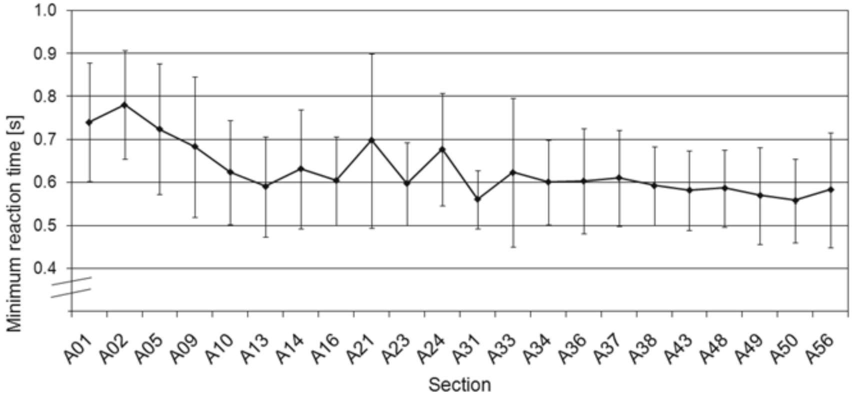


Figure 36. Development of minimum reaction-time values averaged per section along the entire course. Error bars indicate plus / minus one standard deviation.

Nevertheless, an attempt was made to determine the effects of training with the PDT. To do so, the minimum reaction times were compared between the sections with PDT (see Figure 36). Because most sections contained straight road sub-sections the minimum reaction times should reflect performance under comparable low demand. Characteristic differences in reaction time could therefore be interpreted as an effect caused by increasing practice with the PDT.

The pattern of the values depicted in Figure 36 suggested effects of increasing practice with the PDT up to section A13, where a first minimum in RT was reached. Because the values after section A13 were similar to this value, the peaks in sections A21 and A23 are attributed to the specific situational demand in these situations and are not interpreted as the effects of insufficient practice.

Table 33. Comparison of the minimum reaction times in section A13 versus the other PDT-sections: results of t-tests for paired samples.

	Difference				
A13 vs.	*M*	*SD*	*t*	*df*	*p*
A01	-0.15	0.16	-3.63	14	.00
A02	-0.20	0.17	-4.46	13	.00
A05	-0.13	0.14	-3.24	11	.01
A09	-0.09	0.17	-2.09	14	.06
A10	-0.02	0.13	-0.69	13	.51
A14	-0.05	0.08	-2.18	11	.05
A16	-0.01	0.12	-0.31	14	.76
A21	-0.11	0.18	-2.21	11	.05
A23	-0.01	0.12	-0.18	8	.86
A24	-0.09	0.10	-2.83	9	.02
A31	0.01	0.11	0.38	13	.71
A33	-0.02	0.21	-0.28	12	.78
A34	0.00	0.12	0.11	13	.91
A36	0.02	0.12	0.49	12	.64
A37	-0.03	0.14	-0.79	14	.44
A38	0.00	0.12	-0.07	15	.95
A43	0.01	0.10	0.52	14	.61
A48	0.00	0.11	0.14	15	.89
A49	0.01	0.11	0.42	11	.68
A50	0.04	0.12	1.19	14	.26
A56	-0.02	0.15	-0.42	10	.68

The statistical tests in which section A13 was compared to all other sections supported these interpretations. Because the data were normally distributed (non-significant K-S tests), t-tests for paired samples were used for this analysis. Using an ANOVA with all sections was not possible due to the repeated-measures design and the preceding quality assurance process (see above). The combination of both would have resulted in just two participants remaining in the ANOVA sample.

Based on the results shown in Table 33, section A09, which is one of the high accident-rate curves and thus part of the experimental sections, was excluded from further data analyses. Sections prior to section A09 were not part of the experimental sections.

Issues relating to the eye-tracking data

For this study, a contact-free eye-tracking solution was used, which had a number of advantages which were introduced in chapter 4.4.3.9. However, there are some disadvantages owed to the innovative technology used. This means that the technique used for the experiments was not perfect, although it was a high-end solution at this time. This is due to the algorithms used to detect the eyes and the gaze direction. These algorithms have to be applicable to the entire population, which is difficult because of the high variation in facial features. Furthermore, these algorithms and the cameras must be able to track gaze even in constantly changing and often adverse light conditions, as is the case in driving studies in the field. By only selecting participants without glasses, it was at least ensured that no additional source of errors was present in the participant sample. Furthermore, the infrared diodes directed at the drivers' eyes were used in order to compensate for some of the difficulties associated with changing light conditions.

Of course, all providers of contact-free eye-trackers are aware that tracking cannot be perfect all the time. Therefore, the assumed tracking quality is usually recorded in the log-file. In the case of Smart Eye, the tracking quality is given as a value between zero (no tracking) and one (perfect tracking). In theory, data could be selected according to its tracking quality. Such a procedure is described in Weiße (2005) and Weller, Weiße & Schlag (2008). However, for the data at hand, a decision was made against such a pre-selection of data. The argument is based on the same one used against the use of fixations for the spatial distribution of data (see chapter 4.4.3.9): the value which would have to be chosen as the cut-off value would have to be arbitrary, even if well established such as the average plus/minus three standard deviations. Such a cut-off value would not be a problem in a laboratory, where the target area of fixations is usually smaller than the area covered by the cameras. However, in the field study, it is likely that data quality is especially affected when the gaze is directed towards the outer limits of the system or beyond. This was the case each time the PDT was attended to and excluding such data would have probably resulted in an artefact. Therefore, it was appropriate to use all data, unless additional research is conducted which focuses on this very topic in real road driving. Choosing this solution was also possible because the general tracking quality was assured prior to the drives during the intensive calibration process (see chapter 4.4.3.9).

Besides the variability in tracking quality within each experimental drive, the data used here required an additional step with regard to data quality. This was because the two-dimensional coordinates of the gaze data superimposed on the video were not recorded in the data file, but had to be re-calculated by WatchOut from the normalised vector components in three-dimensional space (see chapter 4.4.3.9). While the algorithms used to do this were tested and successfully applied in preceding studies (Lippold & Schulz, 2007; Weichert, 2005), it was necessary to ensure that they still applied to the data set at hand. This is because the experimental car had been used for other studies between those earlier calibration studies and the study used for this thesis, which might have resulted in changes in the hardware configuration. It was also particularly important because the calculation of the gaze parameters was based on these two-dimensional data points (see also chapter 4.4.3.9).

This could be tested quite easily by selecting a single frame from the video where a gaze data point and the frame number were clearly visible and by selecting the matching data entry in the log file. Superimposing the data from the log file on the video frame should result in a match between both data sources.

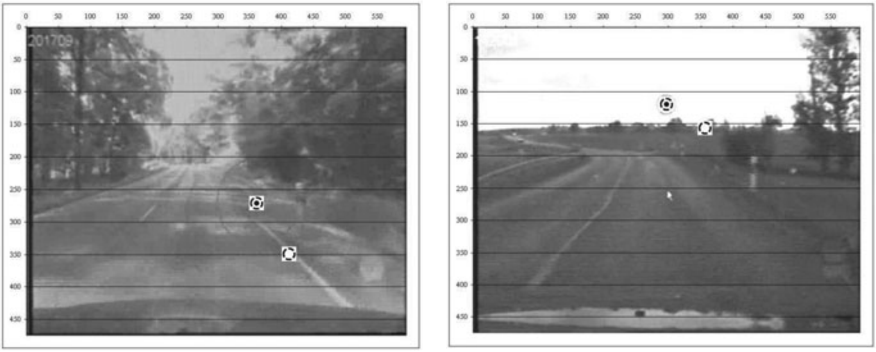


Figure 37. Match between gaze point in the scenery (circle with dot) and calculated gaze point (circle without dot).

As is obvious from Figure 37, the location of the gaze point in the scenery (green cross with red circle) did not match the calculated gaze point (red dot). However, this discrepancy seemed to have been systematic in direction and distance. It also seemed to have been stable for the entire experimental drives because the two pictures in Figure 37 show two different locations (section A37 left and section A14 right) of two different participants (participant No. 06 left, and No. 11 right). Therefore, the origin of the discrepancy might be either a shift of the origin of the Cartesian axis of the calculated data or a shift of the position of the scenery camera used to record the video. Neither of these explanations is very convincing because the algorithms which determine the origin did not change and were tested beforehand, and the change in the camera position should have been noticed during the calibration process. Without an in-depth analysis of this discrepancy, which cannot be conducted as part of this thesis because gaze data was only dealt with in an additional research question, the issue had to remain unsolved.

The question is whether this finding endangers further analysis of the gaze data. This would be the case if it was important whether a certain driver looked at a specific object or not. However, this was not the case because the hypotheses and research questions were not based on such object-based analyses (see chapter 4.4.2), but on parameters of the gaze distribution. A prerequisite of this conclusion is that the error was systematic and present for all subjects and all conditions. Because this was the case (see above), the data could still be used for the analysis.

4.4.3.12 Statistical Methods of Data Analysis

The hypotheses and additional research question 2 required that two or more conditions were tested for differences between them (see chapter 4.4.2 for details). Thus, ANOVAs and t-tests were used, both usually for dependent or paired samples. Prior to their application, data were tested for normal distribution with K-S tests. For additional research question 1, linear regression analysis was used. Because ANOVAs are partly based on linear regression analysis (Field, 2009), the assumptions for ANOVAs also apply to linear regression analysis and were tested accordingly. Finally, the prototypical application of some of the results is shown by using (binary) logistic regression. Unlike the other methods named so far, this is much less common and usually not part of the methodological and statistical education of psychologists. It is therefore explained in more detail for readers interested in understanding the rationale behind this method.

Hypotheses 1 to 5 and additional research question 3

For the statistical analysis of differences between high and low accident-rate curves, t-tests for paired samples were used. Repeated-measures ANOVAs could not be performed because the sample size available at the same time in a single analysis was reduced to only five participants due to the quality assurance process (see chapter 4.4.3.11). A potential accumulation of alpha-errors due to several t-tests (see chapter 4.3.3.15) has to be taken into account when discussing the results. In addition to testing pair-wise differences between one low and one high accident-rate curve, pair-wise t-tests were also conducted within the group of high accident-rate curves (Appendix A 4.5). If high accident-rate curves are compared to the same reference curve, there should be no differences within the high accident-rate group. If there were differences, these were used additionally to discuss the results relevant for the hypotheses.

Additional research question 1

For additional research question 1 ('Are speed and workload related?'), linear regression analysis was used. Details concerning the dependent and independent variables are given in chapter 4.4.4.7 where this additional research question was analysed.

Additional research question 2

For the statistical analysis of research question 2 ('Influence of PDT on behaviour'), two-factor ANOVAs for dependent samples were used separately for each of the dependent parameters. The first factor was the influence of the PDT (with versus without PDT), and the second factor represented the influence of the curves (here: curve A14/A44 versus curve A20/A37). This second factor was added because the curves and the approach zone differed in some geometric aspects. Using this second factor also allowed interpreting the data with respect to the hypotheses and further research questions. Potential differences between the two curves in reaction time were analysed with t-tests for paired samples.

Further analyses

The data used for additional research question 2 served as input for an additional analysis presented in chapter 4.4.4.2. This analysis was conducted to clarify which parameters allow a distinction between two PDT conditions. Binary logistic regression analyses were performed for this purpose. This statistical method is usually applied in clinical studies. Recent publications in traffic-related journals (Gabauer & Gabler, 2008; Gross & Jovanis, 2007; Räsänen, Lajunen, Alticafarbay, & Aydin, 2007; Vollrath, Meilinger, & Krüger, 2002; Walker & Brosnan, 2007) allow the assumption that its application is also on the rise in this field, given the appropriate research questions. For this reason, its statistical background is introduced in an extra chapter.

4.4.3.13 Statistical Background: Binary Logistic Regression

The likelihood that a certain value or values in one or more predictor variables will result in a predefined event or state of the dependent variable can be estimated by applying logistic regression (Backhaus et al., 2006; Garson, 2008; Hosmer & Lemeshow, 2000; Rudolf & Müller, 2004; SPSS 16 Tutorial). In contrast to discriminant analysis, an example of which can be found in Weller et al. (2008), binary logistic regression is a robust method that has no preconditions concerning normally distributed data or similarity of variance/covariance matrices (Backhaus et al., 2006). The implementation of this method in SPSS also allows the validation of the logistic regression function developed for one subset of the data with a second subset of the data. Logistic regression differs in many aspects from linear regression analysis as was used in chapter 4.4.4.7. The main difference is that the dependent variable in logistic regression is not a continuous metric variable, but a nominal variable with different categories. In the case of binary logistic regression, this event is dichotomous, that is, it has just two possible states.

In contrast to a linear regression, the aim of a logistic regression is thus not to approximate the value of the dependent variable but to determine the likelihood of one of the events or states represented by its value. However, likelihoods are defined solely for values between zero and one, whereas a linear regression equation may result in any value within the range minus to plus infinity. In order to transform the outcome of a linear regression equation into the likelihoods needed as result of a logistic regression, the logistic function is used (Backhaus et al., 2006):

$$p_k(y=1)=\frac{1}{1+e^{-z_k}} \tag{25}$$

with:

$$z_k=\beta_0+\sum_{j=1}^{J}\beta_j x_{jk}+u_k \tag{26}$$

where:

$p_k(y=1)$ = probability that the k^{th} case experiences the event y = 1

z_k = value of the unobserved continuous variable resulting from the regression equation

β_0 = constant

β_j = j^{th} regression coefficient

x_{jk} = j^{th} predictor (the independent variable) for the k^{th} case

u_k = residual for the k^{th} case in the regression function.

Conversely, z can be calculated from p as:

$$z_k = \log\left(\frac{p_k}{1-p_k}\right) \tag{27}$$

The regression coefficients are estimated in order to maximise the number of observations that are correctly classified by the regression function. This is done iteratively by maximum likelihood methods with the LogLikelihood function (LL) being used in the logistic regression analysis. The higher the LL-value, the better the result for the classification. By multiplying the LL value with minus two, the resulting new value (-2LL) follows a chi-square distribution with N - J - 1 degrees of freedom (N is the number of participants or observations, and J is the number of dependent variables, Backhaus et al. 2006) and can thus be tested for significance when compared to the corresponding tabulated value of the chi-square distribution. Smaller values of the -2LL-value in relation to the chi-square value indicate good model fit. The measure as such is controversially discussed in the literature (Backhaus et al., 2006) because it depends heavily on the sample size (see the calculation of the df above).

The -2LL-value of the final model with all weighted independent variables can also be compared to the -2LL-value resulting from a regression function in which only the constant is included. In this case, the βs are set to zero. The difference between both -2LL-values is again chi-square distributed with degrees of freedom equalling the number of independent variables. Larger empirical values in comparison to the chi-square value indicate a good model fit. In SPSS 16, the test statistics are provided by the Omnibus-test which tests the null hypothesis that the empirical chi-square value is found without the weighted independent variables. Thus, alpha-error probabilities smaller than .05 or .01 indicate a good model fit.

Other indicators of the quality of the regression function are provided by pseudo-R^2-statistics. They are called pseudo-R^2-statistics because they do not indicate the amount of explained variance as does linear regression. The interpretation of the resulting values with respect to the model quality is nevertheless similar: values closer to one indicate a good model fit. Of the two pseudo-R^2-parameters given in SPSS, Nagelkerke's R^2 is preferred to Cox & Snellen's R^2. This is because, in contrast to Cox & Snellen's R^2, Nagelkerke's R^2 can reach a value of one, which allows a meaningful interpretation. Mathematically this is achieved by dividing the empirically found Cox & Snellen's R^2 by the maximum R^2. This maximum R^2 is the R^2 achieved for the constant-only model.

Finally, the classification results of the regression model itself can be used to determine its quality. In addition to the number of correctly and wrongfully classified cases (i.e. hits versus false positive and false negatives), the proportion of these cases is also used. These proportions are compared to the proportion that can be achieved by chance alone, separated for each of the subsamples. In the case of a dichotomous dependent variable and an equal number of cases in the subgroups, the classification by chance is 50%. In the case of unequal numbers in the subgroups, the hit rate by chance differs depending on which case of the dependent variable is considered. Formulas to calculate these proportional chance rates are given in Backhaus et al. (2006) where formulas are also given for dependent variables with more than two categories. A statistical test based on the classification results is the Hosmer-Lemeshow goodness-of-fit test. This tests the null hypothesis that the difference between actual and predicted classification results is zero. The alpha-error probabilities should therefore be far above .05 in order to indicate a good model fit.

The quality of the regression can be influenced to a large extent by single cases. Identifying such 'critical' cases therefore plays an important role when conducting a logistic regression analysis. These cases are usually characterised by extreme values in the regression function, given their specific values in the independent variables. In order to ensure that the model quality is not influenced by such single cases, a case-wise analysis of the regression results is advised (Backhaus et al., 2006). This is performed by relating the actual classification ($y_k = 0$ or $y_k = 1$ in the case of dichotomous dependent variables) to the likelihood resulting from the regression function for this specific kth case, that is, $p_k(y)$. Mathematically this is expressed by calculating the residual as:

$$RESID_k = y_k - p_k(y) \tag{28}$$

These residuals are usually standardised according to Pearson, whereby RESID-values above 0.5 lead to an increase in the standardised Z-RESID-values. The Z-RESID-values are calculated according to Backhaus et al. (2006) as:

$$ZRESID_k = \frac{y_k - p(y_k = 1)}{\sqrt{p(y_k = 1)(1 - p(y_k = 1))}} \tag{29}$$

with the parameters as explained above.

An example of how these values can be used to identify outliers in the data is shown in Figure 48.

After the model quality has been analysed as described above, the importance of each of the different independent variables in the regression model has to be assessed. This is done by the Wald statistic. The Wald statistic is calculated according to Backhaus et al. (2006) as:

$$W = \left(\frac{b_j}{s_{b_j}} \right)^2 \tag{30}$$

where:

b_j = jth regression coefficient b

s_{b_j} = standard error of the jth regression coefficient b.

W is approximately chi-square distributed with a degree of freedom of one. The test itself tests the null hypothesis that the regression coefficient of the independent variable is zero, which would indicate no influence in the model. Error probabilities less than .05 or .01 can thus be interpreted as significant influence of the independent variable.

After the importance of each independent variable in the regression model has been assessed, the importance of the independent variable with respect to the outcome of the regression must be assessed. In linear regression analysis, one way of achieving this is by comparing the standardised beta weights. In logistic regression, the regression weights cannot be interpreted that straightforwardly [see Formulas (25) and (26)]. Therefore, in logistic regression, the odds ratio is given for each independent variable. An odds is the ratio of the conditional probability of an event occurring and the corresponding probability of this event not occurring, both times given a specific value of the independent variable. The odds that the dependent variable y equals one, given that the independent variable x equals one, is thus calculated as (adapted from Backhaus et al., 2006; Rudolf & Müller, 2004; Hosmer & Lemeshow, 2000; and Field, 2009):

$$\text{Odds}\,(\mathrm{y} = 1) = \frac{p(y = 1 \mid x = 1)}{p(y = 0 \mid x = 1)} = \frac{p(y = 1 \mid x = 1)}{1 - p(y = 1 \mid x = 1)} \tag{31}$$

where:

$p(y = 1 \mid x = 1) = \frac{n_{y=1}}{n_{x=1}}$ = probability that y = 1 given that x = 1

$n_{y=1}$ = number of observations for y = 1

$n_{x=1}$ = number of observations for x = 1.

The odds ratio is the ratio of two odds: firstly the odds that the dependent variable equals one given that the independent variable is also one [see Formula (31)], and secondly the odds that the dependent variables is one given that the independent variable is zero. Thus, the odds ratio (OR) for the event or case y = 1 is defined as:

$$OR = \frac{p(y=1 \mid x=1) / [1 - p(y=1 \mid x=1)]}{p(y=1 \mid x=0) / [1 - p(y=1 \mid x=0)]} \tag{32}$$

The odds and the odds ratio are defined for [0; +∞]. For Formula (32) the OR-value indicates that a participant or case which has a value of x = 1 in the independent variable (one could also say: 'which belongs to group x = 1') is OR-value times more likely to also have a value of y = 1 in the dependent variable (or: 'belongs to group y = 1') than if this participant or case had a value of x = 0 in the independent variable (or: 'belonged to group x = 0'). An OR-value of one indicates that the state of x has no influence on the state of y, that is, whether a participant or case belongs to x = 0 or x = 1 does not influence its chance of belonging to y = 1 or y = 0. A prototypical example of how probabilities, odds, and the odds ratio are calculated is given in Appendix A 1.7 with fictitious data.

If the independent variable is a continuous metric variable (i.e. a covariate), the OR indicates that an increase in this covariate for one unit increases (i.e. OR > 1) or decreases (i.e. OR < 1) the likelihood of the event under consideration. However, this also means that the value of the OR depends on the units used for the independent variable and that OR for dichotomous or categorical (polychotomous) variables differs from continuous variables. In the SPSS-output for the logistic regression, the odds ratio is denoted by the parameter 'Exp(B)'. This is because of the simple relationship between the odds ratio and B which is $OR = e^{\beta_i}$ and which is written as Exp(B). If the units of a continuous independent covariate have to be changed after the calculation of the regression, the OR can be adjusted by multiplying B and using this adjusted B for the calculation of OR. Similarly, the OR can be adjusted for differences between two values (X1, X2) of the covariate as Exp[B × (X2 - X1)] (Hosmer & Lemeshow, 2000).

Applying binary logistic regression to the data at hand has certain shortcomings. Firstly, logistic regression was not designed for dependent samples. Secondly, the number of cases per group should not be smaller than 25 (Backhaus et al., 2006). This is not the case here as the division of the data in subsets results in approximately 15 cases for each condition. However, given the exploratory nature of this analysis in combination with the robustness of the method (see above), it was decided to perform a binary logistic regression nevertheless.

4.4.4 *Results*

The next chapters are generally presented in order of the numbers of the hypotheses. An exception is made for additional research question 2, which is placed at the very beginning of this chapter. This is because its content is an additional test of the sensitivity of the variables and parameters used for the later testing of the hypotheses. The results, therefore, contribute to the later interpretation of the results for the hypotheses.

4.4.4.1 Additional Research Question 2: Influence of the PDT on Behaviour

Despite the rationale of the dual task method (see chapter 4.4.2), the PDT will almost certainly increase the demand of the driving task. Therefore, speed will likely be affected by the PDT, together with speed-related parameters such as acceleration. Attending to the PDT will almost certainly change gaze behaviour (see the '*peripheral*' in PDT). Whether and to what extent this assumption is confirmed by data is the topic of the next paragraphs. Two high accident-rate curves were used to answer additional research question 2:

- curve A14/A44 which was driven with PDT in the outbound direction (A14), and without PDT in the inbound direction (A44), and
- curve A20/A37 which was driven without PDT in the outbound direction (A20), and with PDT in the inbound direction (A37).

Both curves are depicted in Figure 38 for the direction driven without PDT; an impression of the two curves when approached from the opposite direction was already given in Figure 37.

These two curves were particularly appropriate for the question at hand because the preceding road had the same geometric characteristics, regardless of the direction in which the curve was driven. Curve A14/A44 contained a single circular curve with transition curves that was situated between two villages at a distance of approximately 1000 m in both directions from these villages. It was also preceded and followed by a curve with a similar, larger radius of either 270 m or 256 m. The curve direction of these two additional curves is opposite to the curve direction of the main curve (see Figure 38). For the main curve, the curve direction is left for the outbound direction driven with PDT and right for the inbound direction driven without PDT.

Section A20/A37 was also a single circular curve with transition curves. It was situated approximately 500 m away from a village when driven in the outbound direction (A20) and approximately 800 m away from another village when driven in the inbound direction (A37). Curve direction is to the right for the outbound direction driven without PDT and to the left when driven in the inbound direction with PDT. Further details concerning geometry were already given in chapter 4.4.3.3.

Figure 38. Pictures of sections A20 (left) and A44 (right). The latter also shows the preceding curve in opposite direction to the main curve (picture source: RoadView).

The dependent variables and their parameters used to test differences between the sections were those which were also central for the hypotheses and the additional research questions:

- speed (minimum, maximum, average, percentage decrease in speed from maximum speed before the curve to minimum speed in the curve);
- lateral acceleration (maximum);
- longitudinal deceleration (maximum);
- fixation duration (maximum, average);
- number of fixations (maximum, average);
- spatial density index (maximum, average);
- scan path in X- and Y-direction (abbreviated ‘scan path XY’; parameters maximum and average); and
- standard deviation of gaze locations (maximum, average).

The sub-sections which were used for the calculation of the parameters covered a distance from approximately 150 m before the curve up to 50 m after the curve.

Some of the data are depicted in the subsequent Figure 39 and Figure 40 prior to the statistical analyses. These figures depict the values for speed and interpolated reaction time, aggregated for the sub-sections and referenced to road geometry.

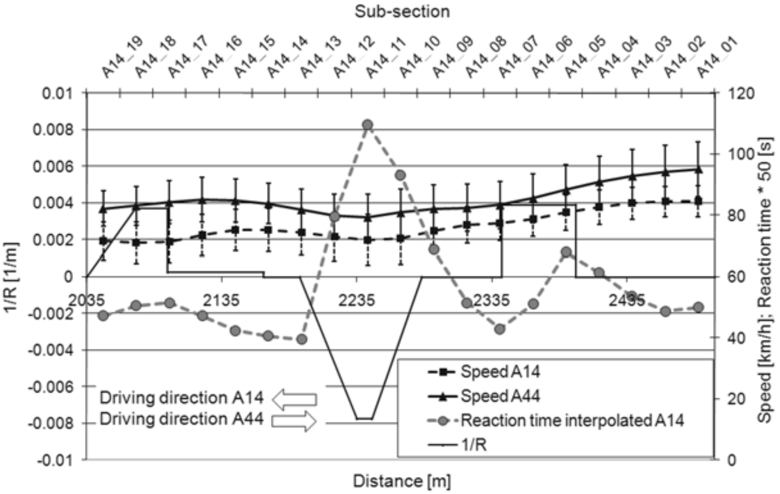


Figure 39. Speed and interpolated reaction time aggregated for the sub-sections of section A14 (with PDT) and A44 (without PDT).

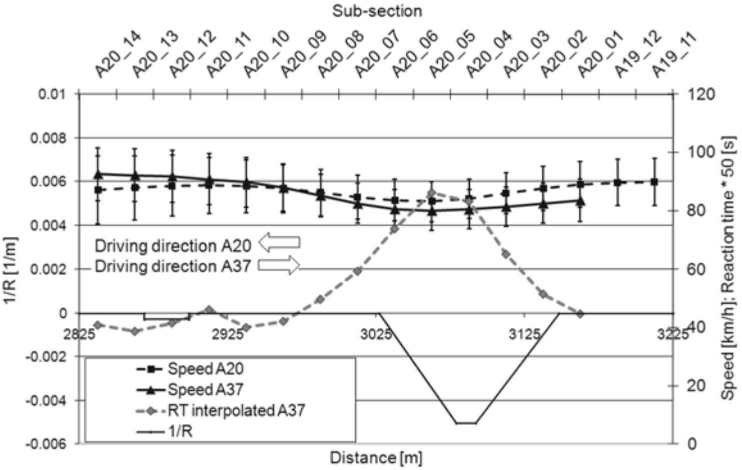


Figure 40. Speed and interpolated reaction time aggregated for the sub-sections of section A37 (with PDT) and A20 (without PDT).

Figure 39 shows curve A14/A44 which was driven with PDT in the outbound direction (A14) and without PDT in the inbound direction (A44). The driving direction for section A14 was from the right to the left in Figure 39, whereas it was from the left to the right for section A44. Figure 40 shows the respective values for the second curve which was used to determine the influence of the PDT on behaviour. Like the preceding curve A14/A44, curve A20/A37 was a left-hand curve when driven with PDT (A37) and a right-hand curve when driven without PDT (A20). The decrease in speed at the end of section A20 (on the left side of Figure 40) was because some participants had already started with the PDT for the next section A21. This did not affect the subsequent statistical analyses because values were aggregated no more than 50 m after the curve and because participants did not know when a new PDT-section would begin.

The comparison between the graphs shown in Figure 39 and Figure 40 already suggested that the influence of the PDT on speed depended on the curve characteristics and thus supported the decision to use the curve as the second factor in the ANOVAs. Figure 41 shows the parameter minimum speed aggregated from the data shown in the two preceding figures.

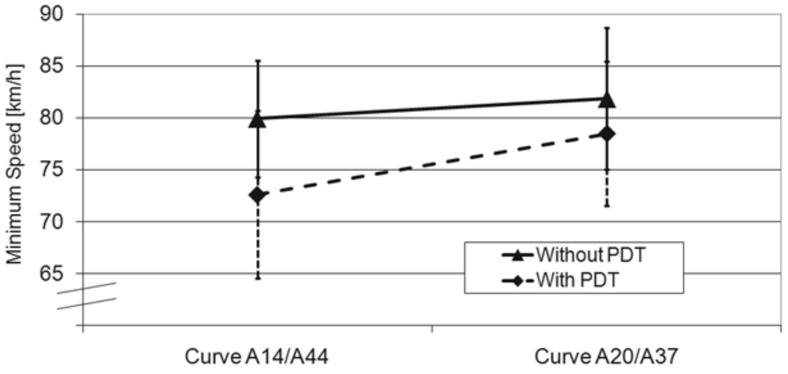


Figure 41. Potential interaction effect between curve characteristics and PDT for the parameter 'minimum speed' and two curves.

The results of the statistical analyses showed that both PDT and curve characteristics (denoted 'curve') had a significant influence on speed (Table 34) and acceleration (Table 36). PDT mainly resulted in a decrease in minimum speed, that is, the speed in the curve itself. Because maximum speed was not affected by the PDT, the decrease in average speed and the increase in the percentage change of speed were a consequence of the decrease in minimum speed. The influence of the curves corresponded to the differences in the geometric curve characteristics: the larger radius and lower CCR-values of curve A20/A37 resulted in a higher speed, lower lateral acceleration and lower longitudinal deceleration (see also Table 36).

Table 34. Effects of the PDT on different parameters of speed in two curves: results of two-factor repeated-measures ANOVAs.

	Descriptive Statistics *M (SD)*		Results for the ANOVAs			
Curve	Without PDT	With PDT	Effect	*F*(1, 9)	*p*	η^2
Minimum speed						
A14/A44	79.92 (5.63)	72.60 (8.06)	Curve	6.08	.04	.40
A20/A37	81.86 (6.86)	78.48 (6.92)	PDT	16.76	.00	.65
			Curve × PDT	1.30	.28	.13
Maximum speed						
A14/A44	87.24(6.68)	84.54 (4.94)	Curve	1.90	.20	.17
A20/A37	87.66 (7.07)	88.87 (8.70)	PDT	0.27	.62	.03
			Curve × PDT	1.03	.34	.10
Average speed						
A14/A44	83.59 (6.12)	77.53 (6.16)	Curve	4.26	.07	.32
A20/A37	84.75 (6.97)	82.96 (7.46)	PDT	10.17	.01	.53
			Curve × PDT	1.51	.25	.14
Percentage change in speed						
A14/A44	8.28 (3.50)	14.28 (5.87)	Curve	4.03	.08	.31
A20/A37	6.63 (1.06)	11.50 (4.99)	PDT	17.99	.00	.67
			Curve × PDT	0.26	.62	.03

The fact that no significant differences were found for maximum speed is also an indication that the two locations were comparable in terms of their approach zones. In combination with the non-significant PDT-effect on maximum speed, this implies even more importantly that the sections before each location were indeed comparable when approached from different directions. Furthermore, the sub-sections chosen for the calculation of maximum speed were sufficient to minimise the influence not only of the curves themselves, but also of potentially irrelevant influences far away from the curves.

Table 35. Correlations and differences in different speed parameters between two curves when driven without and with PDT: results for *t*-tests.

Parameter	*r*	*p (r)*	*t* (*df* = 9)	*p*
Without PDT (curve A44 vs. A20)				
Min.	.32	.37	-0.83	.43
Max.	.18	.62	-0.15	.88
M	.25	.49	-0.46	.66
SD	.44	.20	0.79	.45
% Change	.58	.08	1.73	.12
With PDT (curve A14 vs. A37)				
Min.	.52	.13	-2.49	.03
Max.	.51	.13	-1.83	.10
M	.51	.13	-2.51	.03
SD	.28	.44	-0.03	.98
% Change	.33	.35	1.39	.20

Note. The descriptive statistics are shown in Table 34.

The relationship of the descriptive values for minimum speed (Figure 41) indeed suggests the presence of an interaction effect. However, this effect was not significant for minimum speed or for any other speed parameter (Table 34).

Although this non-significant interaction effect did not indicate a need for further data analyses, additional *t*-tests were calculated. This was because a potential interaction effect would be plausible on theoretical grounds and, more importantly, its presence or absence would be relevant for the interpretation of additional data analyses. In further analysis, it should be ensured that the non-significant interaction effect was not attributed to the small sample size in combination with the rather high variation in the data. Nevertheless, even if this further analysis is in favour of an interaction effect, this effect will remain minor compared to the effects found for curve and PDT.

The results shown in Table 35 can be interpreted in favour of such an assumed interaction effect: differences in speed, which were present but not significant without PDT, were pronounced when the PDT was attended to. While the interaction effect in the ANOVAs was not significant for speed, this effect was significant for lateral acceleration, together with the main effects for curve and PDT (Table 36).

Table 36. Effects of the PDT on maximum lateral acceleration and longitudinal deceleration in two curves: results of two-factor repeated-measures ANOVAs.

	Descriptive Statistics *M (SD)*		Results for the ANOVAs			
Curve	Without PDT	With PDT	Effect	*F*(1, 9)	*p*	η^2
			Maximum lateral acceleration			
A14/A44	4.21 (0.62)	3.13 (0.61)	Curve	11.94	.01	.57
A20/A37	3.28 (0.58)	3.11 (0.54)	PDT	18.49	.00	.67
			Curve × PDT	7.29	.02	.45
			Maximum longitudinal deceleration			
A14/A44	-2.14 (0.92)	-1.76 (0.28)	Curve	10.75	.01	.54
A20/A37	-1.32 (0.23)	-1.52 (0.62)	PDT	0.30	.60	.03
			Curve × PDT	2.15	.18	.19

The values indicate that the reduction in speed caused by the PDT in curve A14/A44 resulted in a similar lateral acceleration as found in curve A20/A37, although the radius of curve A14/A44 was smaller. The higher lateral acceleration in curve A14/A44 without PDT could also mean that participants underestimated this curve in this condition. Adding the PDT might have caused participants to 'give the curve a second thought'.

Of course, with the significant influence of curve characteristics on speed and lateral acceleration the question arose as to whether these differences could also be found in the reaction-time values. Answering this question is also relevant for the testing of the hypotheses and additional research question 1 which deals with the relationship between speed and workload. T-tests for paired samples were calculated for different reaction time parameters (Table 37) instead of ANOVAs because the analysis was necessarily restricted to the conditions driven with PDT. Nevertheless, to ensure comparability with the results for speed, only those participants were selected who could also be used for the ANOVAs with speed shown previously in Table 34.

Table 37. Correlations and differences in different parameters of reaction time between curve A14 and A37.

Parameter (Curve)	*M*	*SD*	*r*	*p* (*r*)	*t* (*df* = 9)	*p*
Max. (A14)	2.85	1.15				
Max. (A37)	2.42	1.63	.17	.64	0.74	.48
M (A14)	1.16	0.15				
M (A37)	1.16	0.47	-.15	.67	-0.04	.97
% Change (A14)	72.19	16.43				
% Change (A37)	63.67	19.03	-.09	.81	1.03	.33

Surprisingly - and strictly speaking contrary to the rationale of the dual task method – the differences in geometry which caused the differences in speed did not affect reaction-time values. This could mean that drivers tried to achieve a homeostatic regulation of workload by diminishing speed with increasing geometric demand. In this case, the PDT itself seemed to have been regarded as the primary task by the drivers and might have been regarded as an additional demand itself (see the significant PDT influence on speed). However, the non-significant results in reaction time should not be stressed too much until further results have been analysed with respect to additional research question 1 (see chapter 4.4.4.7).

Such careful interpretation is also indicated by the non-significant correlation in the reaction-time values and the quite high standard deviations. Of course, the former could also be interpreted in favour of the specific PDT used for this thesis: the task itself was quite easy, which diminished systematic influences of driver characteristics. At the same time it was sensitive enough for very pronounced differences in geometric demand as was the case when straight road sections were compared to curves, which can be seen in the development of the values in Figure 39 and Figure 40.

After the analysis of speed, acceleration and reaction time, the following paragraphs deal with the analysis of differences in gaze behaviour between curves and between the condition with versus without PDT. A first impression of potential influences can be gained from Figure 42 to Figure 45 showing the coordinates of the gaze data points. The direction of the Y-axis is mirror-reversed due to the coordinates used by the system integrated into the experimental vehicle (see Weichert, 2005). In order to increase comparability between all sections, only the ten participants were selected who could be used for all four locations (A14, A44, A20 and A37).

The data shown in Figure 42 to Figure 45 already allowed inferences to be made with regard to the standard deviations of the gaze data. Although the influence of the PDT was significant, the influence of the curve was not significant, nor was an interaction effect curve × PDT (Table 38).

Table 38. Effects of the PDT on maximum and average standard deviation of the gaze locations in two curves: results of two-factor repeated-measures ANOVAs.

	Descriptive Statistics *M (SD)*		Results for the ANOVAs			
Curve	Without PDT	With PDT	Effect	*F*(1, 9)	*p*	η^2
			Maximum SD			
A14/A44	156.4 (98.0)	494.4 (105.7)	Curve	0.54	.48	.06
A20/A37	219.9 (284.4)	505.2 (140.2)	PDT	34.90	.00	.80
			Curve × PDT	0.36	.56	.04
			Average SD			
A14/A44	31.3 (18.9)	284.7 (69.7)	Curve	0.03	.86	.00
A20/A37	44.9 (61.4)	276.9 (82.7)	PDT	124.92	.00	.93
			Curve × PDT	0.25	.63	.03

The next variable of the gaze data which was analysed was the scan path. This variable also takes into account the temporal succession of the gaze data and thus cannot necessarily be derived from the preceding Figures.

Figure 46 gives an impression of this variable by showing its development over distance, separated according to PDT-condition. It is interesting to note that with the beginning of a new PDT section at the end of section A20 (see Figure 40), the scan path increased to the level of section A37 in which the PDT was performed (see left side in Figure 46). As has already been stated, this effect did not influence the values used for the statistical analyses because values were only aggregated until approximately 50 m after the curve. The scan path was much higher when the PDT was attended to, but did not show any significant differences between the two curves (Table 39).

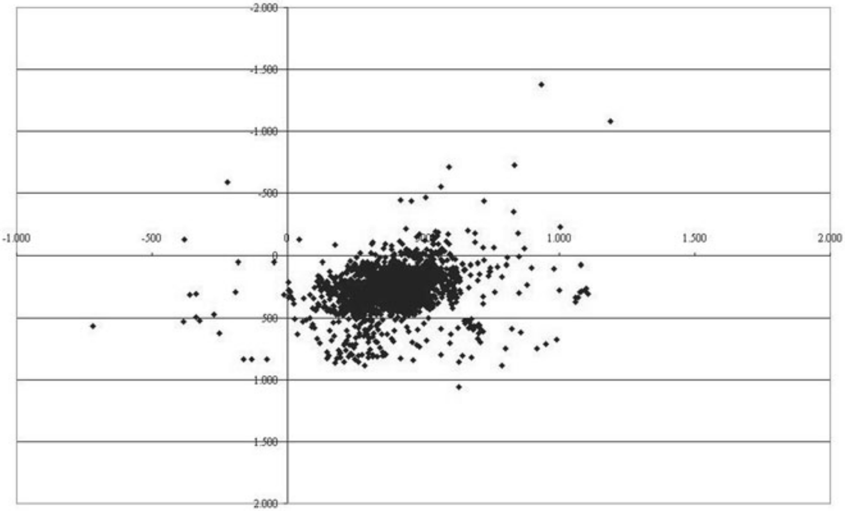


Figure 42. Distribution of gaze data for ten participants and curve A44 driven without PDT [pixel].

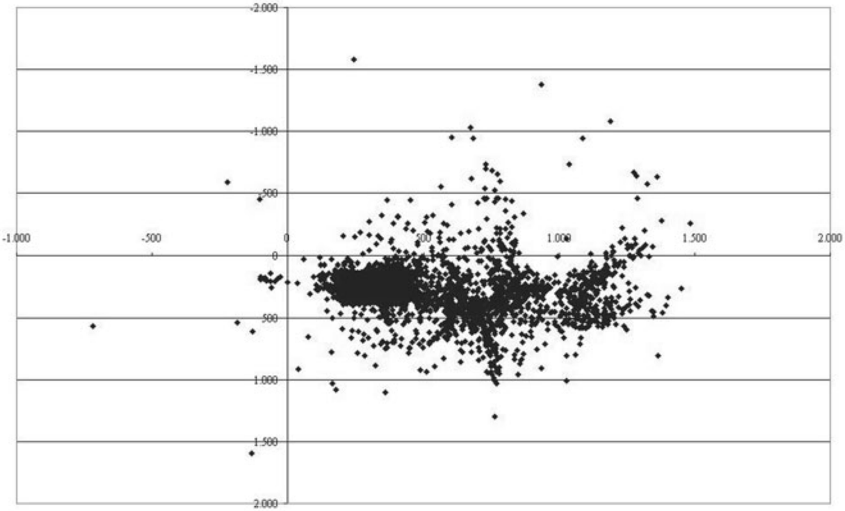


Figure 43. Distribution of gaze data for ten participants and curve A14 driven with PDT [pixel].

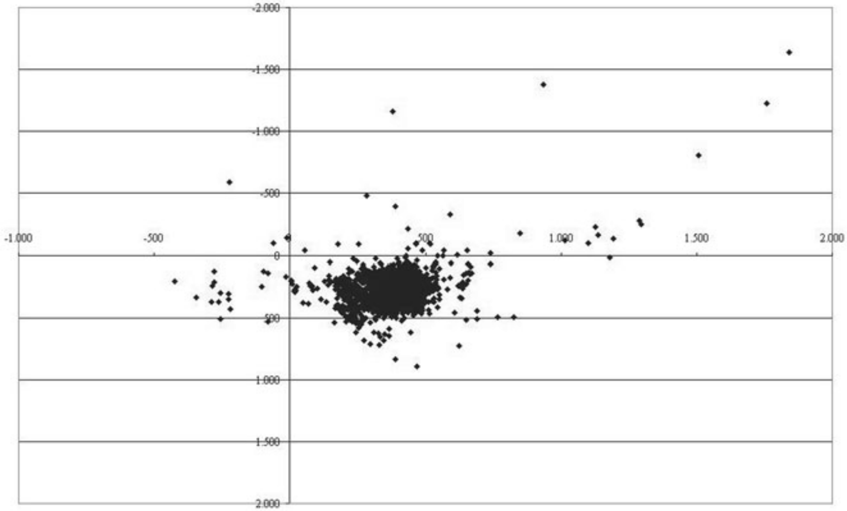


Figure 44. Distribution of gaze data for ten participants and curve A20 driven without PDT [pixel].

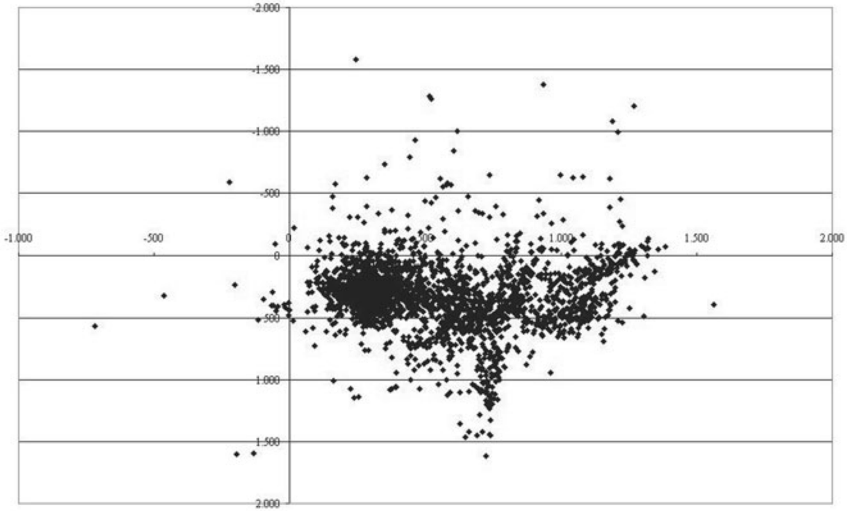


Figure 45. Distribution of gaze data for ten participants and curve A37 driven with PDT [pixel].

Table 40 shows the results for the spatial density index (SDI). Again the results showed a highly significant influence of the PDT with no differences between the two curves except for the parameter maximum SDI. The lower SDI-values for the conditions with PDT meant that fewer fields had been ‘looked at’ (see chapter 4.4.3.9 for details). This is somewhat surprising because the gaze data shown in the figures above implies that more fields were looked at in the PDT-condition.

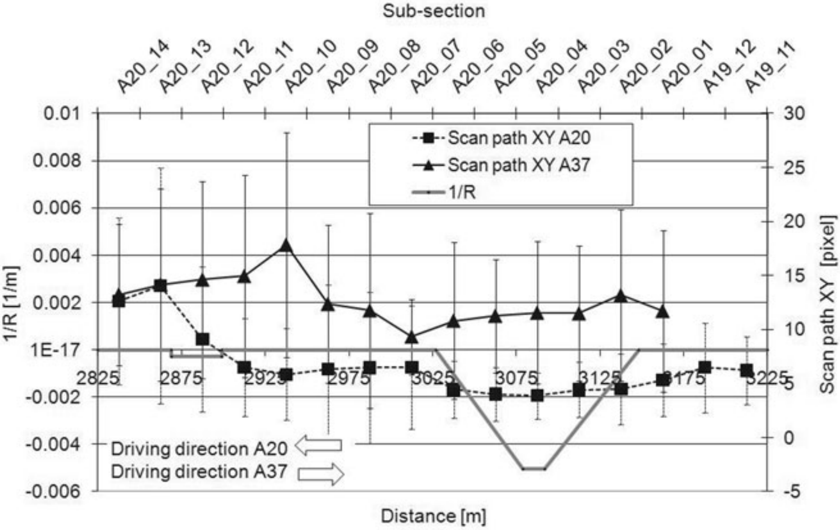


Figure 46. Scan path in XY-direction combined, aggregated for the sub-sections of section A37 (with PDT) and A20 (without PDT).

The reason for the lower spatial density index is that only the area covered by the scenery camera was used for the calculation of the spatial density index (see also chapter 4.4.3.9). Because more fields were looked at outside this area when the PDT was attended to, fewer fields were looked at inside the area covered by the scenery camera, hence the lower spatial density index in the PDT condition. The area covered by the scenery camera was situated below the X-axis in Figure 42 to Figure 45. It was extended to 594 pixels in the X-direction and to minus 475 pixels in the Y-direction (Weichert, 2005). This area covered the road ahead and thus was the centre of the gaze data.

Table 39. Effects of the PDT on the scan path in X- and Y-direction in two curves: results of two-factor repeated-measures ANOVAs.

	Descriptive Statistics *M (SD)*		Results for the ANOVAs			
Curve	Without PDT	With PDT	Effect	*F*(1, 9)	*p*	η^2
			Maximum scan path			
A14/A44	15.66 (5.68)	24.94 (9.44)	Curve	0.01	.95	.00
A20/A37	13.16 (7.98)	27.72 (7.62)	PDT	25.44	.00	.74
			Curve × PDT	1.13	.32	.11
			Average scan path			
A14/A44	7.68 (3.28)	11.93 (3.83)	Curve	0.00	.98	.00
A20/A37	5.93 (2.51)	13.73 (4.10)	PDT	30.14	.00	.77
			Curve × PDT	2.13	.18	.19

Table 40. Effects of the PDT on the SDI in two curves: results of two-factor repeated-measures ANOVAs.

	Descriptive Statistics *M (SD)*		Results for the ANOVAs			
Curve	Without PDT	With PDT	Effect	*F*(1, 9)	*p*	η^2
			Maximum spatial density index			
A14/A44	0.030 (0.009)	0.026 (0.010)	Curve	0.13	.73	.01
A20/A37	0.032 (0.006)	0.022 (0.007)	PDT	33.08	.00	.79
			Curve × PDT	1.90	.20	.17
			Average spatial density index			
A14/A44	0.014 (0.007)	0.011 (0.006)	Curve	0.00	.97	.00
A20/A37	0.017 (0.005)	0.008 (0.005)	PDT	17.27	.00	.66
			Curve × PDT	5.05	.05	.36

Finally, fixations were analysed in terms of both their number and their duration. How these values were defined was explained in chapter 4.4.3.9. The development of fixation duration with distance is shown for curve A14/A44 in Figure 47.

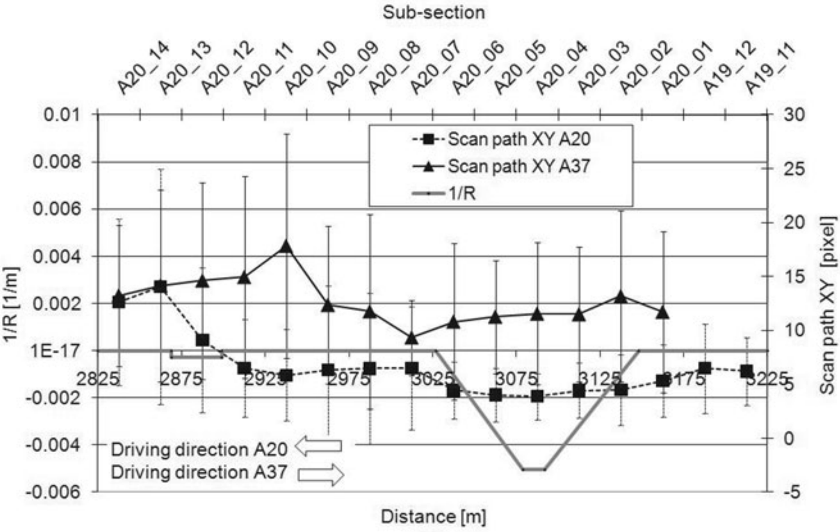


Figure 47. Averaged fixation duration for each sub-section of section A14 (with PDT) and A44 (without PDT). Error bars indicate plus/minus one *SD*, driving direction for A14 was from right to left and vice versa for A44.

It is interesting to note that there was a considerable increase in fixation durations at the beginning of the curve, independent of direction or PDT condition. This indicates an increased need for information at this point. It was also mirrored in the increase in reaction times after this increase in fixation duration (ongoing reaction process). This effect was not analysed further, but is an interesting aspect for future studies. The statistical analyses of curve effect and PDT-influence are shown in Table 41 and Table 42.

The influence of the PDT was significant for the preceding variables, but no significant differences were found between the two curves. However, the direction of the differences caused by the PDT requires further discussion. This is because the results at first sight contradict theoretical assumptions regarding the relationship between fixation duration and number of fixations: within a given period of time, an increase in fixation duration would usually be assumed to result in a decrease in the number of fixations (Weiße, 2005; Weller, Weiße et al., 2008). Contrary to these theoretical assumptions, it was found that the PDT caused shorter fixation durations *and at the same time* fewer fixations. This can be explained with the longer scan paths in the PDT condition (see Table 39). The time needed for these longer scan paths diminished the data which remained available to calculate fixations. This effect might also have interacted with lower data-quality when the PDT was being attended to.

Table 41. Effects of the PDT on maximum and average fixation duration in two curves.

	Descriptive Statistics *M (SD)*		Results for the ANOVAs			
Curve	Without PDT	With PDT	Effect	*F*(1, 9)	*p*	η^2
			Maximum fixation duration			
A14/A44	0.52 (0.28)	0.47 (0.28)	Curve	2.79	.13	.24
A20/A37	0.46 (0.11)	0.31 (0.10)	PDT	5.09	.05	.36
			Curve × PDT	0.48	.50	.05
			Average fixation duration			
A14/A44	0.20 (0.11)	0.18 (0.11)	Curve	1.95	.20	.18
A20/A37	0.20 (0.07)	0.12 (0.07)	PDT	5.03	.05	.36
			Curve × PDT	2.43	.15	.21

In addition to the results for the effect of the PDT, the results were also inconsistent with respect to theoretical models for the effect of the curves. Because speed was found to be higher in section A20/A37 (Table 34), the time available for fixations decreased and therefore, fewer fixations should have been found, given that fixation durations remained the same. Although fixation durations did not differ, the average number of fixations was even higher for curve A20/A37. To analyse which condition (with PDT or without PDT) caused the significant result for the factor 'curves', additional t-tests for paired samples were calculated (Appendix A 4.6). These tests showed that curve A20/A37 had significantly more fixations (both maximum and average) in the condition without PDT, while no differences were found in the condition with PDT. In contrast to the effect for the PDT, the results cannot be explained by differences in the scan path or other variables of gaze behaviour. This is because no other variable exhibited a corresponding pattern of results.

Table 42. Effects of the PDT on maximum and average number of fixations per subsection in two curves.

	Descriptive Statistics M (SD)		Results for the ANOVAs			
Curve	Without PDT	With PDT	Effect	$F(1, 9)$	p	η^2
	Maximum number of fixations					
A14/A44	2.80 (0.79)	2.50 (0.97)	Curve	0.38	.56	.04
A20/A37	3.30 (0.67)	2.20 (0.79)	PDT	16.96	.00	.65
			Curve × PDT	2.44	.15	.21
	Average number of fixations					
A14/A44	1.29 (0.57)	0.97 (0.45)	Curve	5.78	.04	.39
A20/A37	1.81 (0.73)	0.85 (0.45)	PDT	14.69	.00	.62
			Curve × PDT	5.98	.04	.40

Thus, it is likely that differences in the two curves such as for sight distance (clearly visible in Figure 38) did indeed cause differences in gaze behaviour. However, these were not pronounced enough to have caused significant differences. These non-significant differences might have interacted with the temporal or spatial limits used in the definition of fixations (see chapter 4.4.3.9) or the division of the section into subsections. While these interactions in themselves would be an interesting topic for further research together with the supposed decrease in data quality with PDT, they are not relevant for this thesis because all curves used for further analyses in relation to the hypotheses and the research questions were driven with PDT.

Summing up the results regarding gaze data, it was shown that attending to the PDT had such a strong influence on gaze behaviour that differences which might have been present in the condition without PDT disappeared once the PDT was attended to.

4.4.4.2 Prototypical Application for the Functionality of IVIS & ADAS

In the preceding chapter it was shown that the PDT resulted in changes of behaviour. Knowledge regarding these changes is not only important to correctly interpret the results of the hypotheses. The results of the preceding chapter can also be used for practical applications. In fact, the PDT can be seen as a typical secondary task that captures the drivers' attention or to which the drivers wilfully divert their attention. Such distraction or inattention is regarded as the sole or most important contributing factor to accidents (McEvoy, Stevenson, & Woodward, 2007). Detecting such periods of distraction or inattention could significantly reduce these accidents. This functionality does not have to be designed as an additional alerting device for the driver but might be used as an alerting device for the vehicle instead. This could be achieved by a higher technical preparedness to brake whenever such a condition is detected.

Based on the results in the preceding chapter, a variable had to be selected which showed highly significant differences between the condition with and without PDT. Furthermore, in order to ensure a reliable distinction between the two conditions, an influence from environmental characteristics should not be present. This was the case for the variable 'standard deviation of gaze location'. For this variable, the parameter 'average across the entire section' showed the highest eta-square values (Table 38).

Statistically speaking, a binary logistic regression is the method of choice for the task at hand (see chapter 4.4.3.12). For use in a potential alerting system for driver inattention, the regression should distinguish between a condition in which the driver pays attention to the road and a condition in which the driver does not.

Table 43. Descriptive statistics for the parameter *SD*-gaze divided for the conditions and datasets used.

Curve	Parameter	Without PDT	With PDT	Both
A14/A44	n	14	12	26
	M	35.53	304.54	159.69
	SD	18.13	83.76	148.20
A20/A37	n	15	15	30
	M	38.67	276.37	157.52
	SD	51.12	105.56	145.79
Both	n	29	27	56
	M	37.15	288.89	158.53
	SD	38.23	95.77	145.57

For the data used in this thesis, these two conditions are prototypically represented by the two conditions with or without PDT (coded 0 for the condition without PDT and 1 for the condition with PDT). The average standard deviation of the gaze location represented the sole covariate. In addition, a constant was included. The data were divided into two data subsets that were used for the development and validation of the model. The values for curve A20/A37 were selected for the calculation of the model, whereby the values for curve A14/A44 served as input for the validation of the model developed with the data for curve A20/A37. Because there was only one independent variable, it was necessary to calculate the model itself with the inclusion method. Table 43 shows the descriptive statistics for the datasets used.

The different statistics of the logistic regression indicated very good model fit. First of all, the difference in the LL-value between the model without *SD*-gaze and with *SD*-gaze was significant (Omnibus test statistics: $\chi^2 = 29.41$, $df = 1$, $p < = .001$). Secondly, the null hypothesis of the Hosmer-Lemeshow goodness-of-fit test could not be rejected, indicating that the assumed model indeed fits the data ($\chi^2 = 7{,}59$, $df = 8$, $p = .47$). Thirdly, the pseudo-R^2 statistics supported the quality of the regression function with Cox & Snell's pseudo-R^2 being .63 and Nagelkerke's pseudo-R^2 being .83 (the -2LL-value was 12.18). Concerning the number of correctly classified cases, two cases were wrongly classified in the first dataset used to develop the function: one participant was wrongfully assigned to the PDT-condition and one participant was wrongfully assigned to the non-PDT-condition. This resulted in a correct classification of cases of 93.3% and 100% for the subset of data used to validate the function developed with the first dataset. Before analysing these two wrongfully classified cases in more detail, Table 44 shows the statistics for the covariate and the constant.

Table 44. Wald statistics and odds ratios for the standard deviation of the gaze data and the constant in the equation of the logistic regression.

Variable	*B*	*SE*	Wald (*df* = 1)	*p*	OR	CI 95%	
SD-gaze	0.03	0.01	8.73	.00	1.03	1.01	1.05
Constant	-3.62	1.29	7.80	.01	0.03		

With the results given in Table 44 it was not surprising that the Wald statistics were highly significant, indicating the importance of both the variable *SD*-gaze and the constant. The seemingly low value of the odds ratio for *SD*-gaze must be related to the units used. For the data at hand, these units were pixels. Applied to the odds ratio, this means that with every additional pixel in *SD*-gaze, the likelihood that the data belonged to the condition with PDT increased by 1.03. If the difference between the condition without PDT and the condition with PDT were just one pixel, this odds ratio would be meaningless. However, with the approximate average difference between both conditions having been around 200 pixels (Table 43), the finding is indeed meaningful, even at this stage.

As was described in chapter 4.4.3.12, the quality of a logistic regression can be largely influenced by single cases. In fact, the two wrongfully classified cases for the data subset of curve A20/A37 in combination with the large differences in the conditions (Table 43) suggest such a case-wise analysis. Similar to the description in chapter 4.4.3.12, the standardised residuals (*Z*-Resid) were used for this purpose. These values are shown together with the participant-wise *SD*-gaze values in Figure 48.

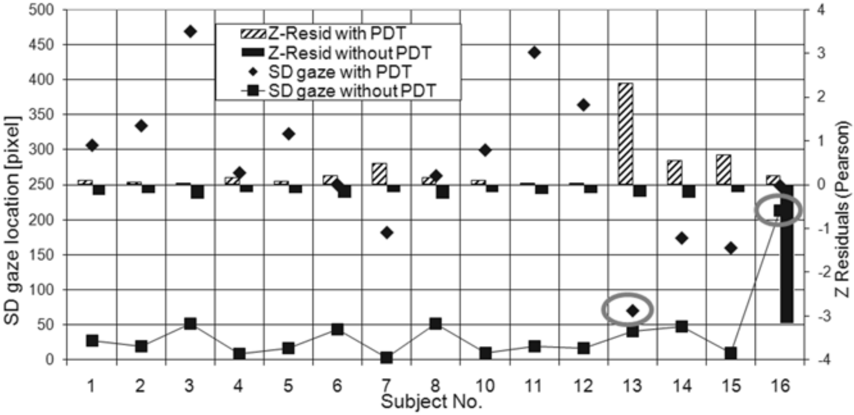


Figure 48. Z-Residuals for the binary logistic regression combined with the standard deviation (SD) of the gaze data to identify outliers.

In Figure 48 the two outliers in the data which represent the two misclassified cases can clearly be identified from the Z-Resid values: participant No. 13 who was wrongfully classified as driving in the non-PDT condition, and participant No. 16 who was wrongfully classified as driving in the PDT-condition. Because *SD*-gaze was used as the single predictor in the logistic regression, the outliers in the *Z*-Resid data are matched by outliers in the *SD*-gaze data, highlighted by the circles in Figure 48.

No explanation could be found for these outliers despite a secondary analysis of the videos and the reaction times. With regard to the videos it must be noted that they were taken from the scenery camera which did not include the driver's head. Therefore, despite analysis of the video, it cannot be ruled out that participant No. 16

was engaged in a behaviour which resulted in the high SD-gaze values in the non-PDT condition.

Although it is not advisable to exclude single cases without proper explanation (Backhaus et al., 2006), this step was nevertheless taken here. This was done because it is assumed that proper explanations can be found for the outliers, although this was not possible for this data because of technical reasons (for example, no video of the drivers' heads). Not surprisingly, the quality of the logistic regression could be further increased by this step: it resulted in 100% of correctly classified cases for both subsets of data (Nagelkerkes pseudo-$R^2 = 1$; Omnibus test statistics: $\chi^2 = 36.00$, $df = 1$, $p < = .001$, OR per pixel = 1.38).

Taken together, these results are extremely encouraging regarding the development of an in-vehicle device to detect driver inattention or distraction. To circumvent the problem of unsatisfying explanations for outliers, future studies must place additional emphasis on data quality and must also record the driver's head.

4.4.4.3 Summary and Discussion of Results Concerning the Influence of the PDT

The influence of the PDT was analysed by comparing behaviour in two curves which were both driven with and without PDT. Knowing whether and how the PDT influenced behaviour without PDT is seen as prerequisite for the meaningful interpretation of the further data analyses. The results can also be interpreted in relation to existing theories of driving behaviour. An additional emphasis was put on the analysis of gaze behaviour, which thus also allowed statements concerning data quality and the quality of the algorithms used for gaze data analysis.

The results showed a high influence of the PDT on 'normal' behaviour. Perhaps the most salient result for this thesis was the effect of the PDT on speed. Here, it was found that the PDT resulted in a decrease in speed in the curve, whereas maximum speed before the curve was not affected. Differences in speed in the curves (i.e. minimum speed) were also found between the two curves. The differences in speed between the two curves mirrored differences in assumed demand caused by the differences in radius and other geometric parameters. The fact that maximum speed before the curves did not differ between curves or between PDT-conditions showed the general suitability of the selected curves for the question at hand.

Although interaction effects of curve and PDT were not significant for speed, additional analyses showed that the PDT tended to increase differences between curves: in the PDT condition, speed disproportionally decreased with higher geometric demand. This interpretation is supported by the different effect the PDT had on speed on the straight road section (i.e. maximum speed) compared to the effect it had in the curve (i.e. minimum speed). This differentiated effect of the PDT on speed also resulted in a levelling effect of lateral acceleration, which should be kept in mind for further analysis of the data.

Given the obvious regulation of speed with demand (see above), the question arose as to how reaction times in the PDT would develop. In fact, no differences in reaction time were found between the two curves when they were driven with PDT. Obviously, the drivers did indeed regulate speed in order to keep workload – as represented by reaction time – within a certain range, independent of geometric demand. This is in fact what would have been predicted by workload homeostasis theory (Fuller, 2005; Gstalter & Fastenmeier, 1995). However, this preliminary interpretation is based on the comparison between curves, not on a comparison within the curves themselves and their preceding road sections. Here, it became obvious that reaction times increased sharply in the curves compared to the preceding straight road sections (see Figure 39 and Figure 40).

Thus, even without statistical analysis one can state that a homeostatic regulation of workload is not achieved when the driving task changes fundamentally, as is the case when driving a straight road section in comparison to a curve. Compared to such fundamental differences, the differences in demand caused by the differences in the geometric characteristics of the curves themselves could be regarded as minor. However, a final interpretation regarding the effect of curve demand itself on the regulation of workload should be performed by including a higher number of curves which was done in chapter 4.4.4.7. Concerning the PDT itself, the results so far suggest that the PDT competes for the same resources as the driving task. This was shown by the trade-off between minimum speed and PDT, and thus also indicates that the PDT used here was a valid method to assess workload in driving (Brown, 1978; Tsang & Vidulich, 2006).

In addition to the variables discussed above, gaze data was analysed. The results differed with respect to the variables used. In general, the results were of higher quality with respect to theoretical assumptions when the raw data and parameters directly derived from the raw data were used. This was the case for the standard deviation of the gaze data and the scan paths, which were both higher when the PDT was attended to. These results were in line with theoretical assumption (Weiße, 2005; Weller, Weiße et al., 2008) and could in fact be used for future driver assistance or information systems (IVIS & ADAS) by reliably distinguishing conditions with or without attending to a secondary task.

However, even for these variables the influence of the PDT was so strong that potential differences between the curves were completely levelled out. It must be noted that such differences were not found in the condition without PDT either. If they had been present there, it should have been indicated at least by a significant interaction effect. The absence of significant differences is even more astonishing because the two curves used differed considerably with respect to their environment (see Figure 38).

The results for gaze variables which required further processing steps were even more difficult to interpret. This was particularly the case for different parameters of fixations. As was described in chapter 4.4.3.9, the definition of a fixation requires temporal and spatial limits which could have influenced the results. The same applied to the spatial density index in the way it was calculated here (see also chapter 4.4.3.9). It might be the case that the results for these processed gaze variables could be improved by improving the algorithms used. However, considerable effort has already been invested in doing so (Lippold & Schulz, 2007; Weichert, 2005) and if such a step was done additionally, a larger sample would be required for an additional experimental setup in the field during driving. This could not be done as part of this thesis, especially as analysing the gaze data is merely a side aspect.

These shortcomings in combination with the fact that all curves selected for the testing of the hypotheses were driven with PDT suggest that there is no point in further analysing gaze data. To do so would only be meaningful without PDT, with a larger sample size, and with additional effort invested in the defining of the algorithms used. Accordingly, it was decided not to further analyse additional research question 3 ('Differences in gaze behaviour between high and low accident-rate curves', see chapter 4.4.2).

4.4.4.4 Differences in Behaviour Between High and Low Accident-Rate Curves (Hypotheses 1 to 5)

The results reported in this chapter concern the comparison between high and low accident-rate curves characterised by similar geometry. Several hypotheses were formulated regarding differences in driver and driving behaviour between these pairs of curves. It was assumed (see chapter 4.4.2 for details) that high accident-rate curves in comparison to low accident-rate curves are characterised by

- a higher maximum speed before the curve;
- a higher maximum longitudinal deceleration;
- a higher percentage decrease in speed;
- a higher minimum speed in the curve;
- a correspondingly higher maximum lateral acceleration;
- a higher workload; and
- a higher percentage change between minimum and maximum workload.

Instead of separating the results, all hypotheses regarding the above-mentioned differences in behaviour have been combined in this single chapter. This decision is based upon the results in the preceding chapters. It was shown there that speed and speed-related parameters have to be analysed conjointly with the results for the PDT. Five pair-wise comparisons between sections including a singular curve (radius given in brackets below) were used for the analyses:

- section A14 (R = 129) versus section A21 (R = 135);
- section A24 (R = 141) versus section A21 (R = 135);
- section A37 (R = 198) versus section A21 (R = 135);
- section A33 (R = 141) versus section A36 (R = 135); and
- section A50 (R = 128) versus section A36 (R = 135).

The first sections in the pairs were the high accident-rate curves, the second ones were the reference curves. These sections have already been described in more detail in chapter 4.4.3.3 together with methodological implications. The data was successfully tested for normal distribution. The results of the t-tests for paired samples (see chapter 4.4.3.12) are given in the subsequent tables and are summarised below. Results for differences within the group of high accident-rate curves are given in Appendix A 4.5. An overall discussion and interpretation of the results is given in the next chapter.

Table 45. Correlations and differences in maximum speed before the curve between high accident-rate curves (first in pairs) and low accident-rate reference curves: *t*-tests for paired samples.

Pairs	*M*	*SD*	*r*	*p (r)*	*t*	*df*	*p*
A14	84.69	6.64					
A21	84.12	9.07	.78	.01	0.32	9	.76
A24	79.65	7.86					
A21	82.74	9.33	.96	.00	3.15	7	.02
A37	92.19	10.16					
A21	86.75	10.30	.88	.00	-3.75	11	.00
A33	70.53	5.86					
A36	85.62	8.14	.74	.01	-9.53	11	.00
A50	88.18	7.25					
A36	85.62	8.14	.58	.05	1.24	11	.24

Table 46. Correlations and differences in percentage change between the maximum and minimum speed between high accident-rate curves (first in pairs) and low accident-rate reference curves.

Pairs	*M*	*SD*	*r*	*p (r)*	*t*	*df*	*p*
A14	15.16	6.26					
A21	15.80	5.46	.48	.16	-0.34	9	.74
A24	15.55	5.66					
A21	16.33	5.04	.51	.20	0.42	7	.69
A37	12.79	6.39					
A21	17.39	6.69	.62	.03	2.79	11	.02
A33	8.07	4.02					
A36	11.46	5.39	.05	.89	-1.79	11	.10
A50	21.30	8.14					
A36	11.46	5.39	.32	.32	4.14	11	.00

Table 47. Correlations and differences in minimum speed between high accident-rate curves (first in pairs) and low accident-rate reference curves (second in pairs): *t*-tests for paired samples.

Pairs	*M*	*SD*	*r*	*p (r)*	*t*	*df*	*p*
A14	71.89	7.94					
A21	70.90	9.34	.67	.04	0.44	9	.67
A24	67.18	7.03					
A21	69.33	9.63	.85	.01	1.17	7	.28
A37	80.04	7.26					
A21	71.49	8.77	.77	.00	-5.27	11	.00
A33	64.77	5.26					
A36	75.66	7.23	.73	.01	-7.68	11	.00
A50	69.16	6.88					
A36	75.66	7.23	.64	.03	-3.77	11	.00

Table 48. Correlations and differences in maximum longitudinal deceleration between high accident-rate curves (first in pairs) and low accident-rate reference curves: *t*-tests for paired samples.

Pairs	*M*	*SD*	*r*	*p (r)*	*t*	*df*	*p*
A14	-1.76	0.23					
A21	-1.82	0.86	.49	.15	0.25	9	.81
A24	-1.68	0.54					
A21	-1.69	0.73	.59	.12	-0.06	7	.96
A37	-1.62	0.64					
A21	-1.89	0.79	.64	.02	-1.53	11	.15
A33	-0.68	0.44					
A36	-1.33	0.62	.38	.23	3.71	11	.00
A50	-1.72	0.60					
A36	-1.33	0.62	.15	.64	-1.72	11	.11

Table 49. Correlations and differences in maximum lateral acceleration between high accident-rate curves (first in pairs) and low accident-rate reference curves: *t*-tests for paired samples.

Pairs	*M*	*SD*	*r*	*p (r)*	*t*	*df*	*p*
A14	3.05	0.46					
A21	2.91	0.63	.35	.32	0.70	9	.50
A24	2.98	0.51					
A21	2.84	0.68	.48	.23	-0.62	7	.55
A37	3.14	0.64					
A21	2.92	0.60	.71	.01	-1.65	11	.13
A33	2.88	0.48					
A36	2.86	0.52	.32	.32	0.15	11	.89
A50	3.17	0.47					
A36	2.86	0.52	.25	.44	1.76	11	.11

Table 50. Correlations and differences in maximum reaction time between high accident-rate curves (first in pairs) and low accident-rate reference curves (second in pairs): *t*-tests for paired samples.

Pairs	*M*	*SD*	*r*	*p (r)*	*t*	*df*	*p*
A14	2.69	1.32					
A21	2.78	1.17	-.06	.87	-0.16	9	.88
A24	2.26	0.99					
A21	2.76	1.34	.13	.75	0.92	7	.39
A37	2.28	1.55					
A21	2.51	1.24	.32	.30	0.47	11	.65
A33	1.00	0.50					
A36	1.41	0.91	.13	.70	-1.41	10	.19
A50	1.66	0.88					
A36	1.43	0.87	-.05	.88	0.62	11	.55

Table 51. Correlations and differences in percentage change between the maximum and minimum reaction time between high accident-rate curves (first in pairs) and low accident-rate reference curves (second in pairs): *t*-tests for paired samples.

Pairs	*M*	*SD*	*r*	*p (r)*	*t*	*df*	*p*
A14	68.36	20.98					
A21	69.79	15.64	.28	.43	-0.20	9	.85
A24	64.26	16.50					
A21	67.79	16.67	.16	.71	0.46	7	.66
A37	63.52	18.85					
A21	65.15	18.14	.13	.69	0.23	11	.82
A33	32.21	16.53					
A36	46.76	20.16	.26	.44	-2.14	10	.06
A50	58.72	17.38					
A36	48.61	20.25	.11	.72	1.39	11	.19

4.4.4.5 Hypotheses 1 to 5: Summary and Discussion of Results

Before the curve-wise discussion of the results, they are summarised with respect to the different variables and parameters. In general, very few of the differences between high and low accident-rate curves were significant. The number of significant results also differed between the variables and parameters: for maximum and minimum speed, three out of five differences were significant, whereas none were significant for maximum reaction time. Results for other parameters and variables were situated between these two extremes. Even for speed, the significant differences were not all in the same direction, which also means that they were not all in the direction assumed in the hypotheses. The results described so far can also be seen as supporting the assumption in previous chapters where it was stated that speed is used to maintain a certain preferred level of workload.

Differences between variables and parameters were also found in terms of their stability across participants and situations. This stability is indicated by the correlation coefficient and its significance. Similar to the results for the t-tests, the highest correlation coefficients and the highest number of significant coefficients were found for maximum and minimum speed, whereas only low coefficients and non significant ones were found for maximum reaction time.

For the subsequent curve-wise discussion of the results, another source of information is used in addition to the comparisons shown in the preceding chapter. These are the results of the comparisons between the curves within the group of high accident-rate curves themselves. Because these results are only necessary for the interpretation of the results regarding the hypotheses, but do not constitute important results per se, they are shown in Appendix A 4.5 rather than in the preceding chapter.

These results (Appendix A 4.5) showed consistent differences for curve A33 when compared to other high accident-rate curves. Curve A33 was characterised by a significant lower maximum speed before the curve, lower minimum speed in the curve, lower percentage change of speed, and a lower maximum longitudinal deceleration. These findings were not surprising and can be explained as a consequence of the preceding village and the speed camera situated in this village (see chapter 4.4.3.3).

Similar results were found for the comparison between the high accident-rate curve A33 and its low accident-rate reference curve A36. Here, maximum and minimum speed were also lower in the high accident-rate curve A33 (Table 45 and Table 47). Because the reference curve was also preceded by a village but without a speed camera (see chapter 4.4.3.3), the latter differences in speed can be attributed to the effect of the speed camera alone. Regardless of their cause, the lower values for speed and the lower value for longitudinal deceleration (Table 48) for the high accident-rate curve A33 in comparison to the low accident-rate reference curve were not in line with the assumptions formulated in the hypotheses.

The reaction times for curve A33 showed similar effects as for speed. It was found that the maximum reaction time in curve A33 and the percentage change between maximum and minimum reaction time were significantly lower in comparison to all other high accident-rate curves (Appendix A 4.5). The results in comparison to the low accident-rate reference curve A36 were less pronounced but pointed in the same direction. The maximum reaction times were also lower in the high accident-rate curve A33, although this difference was not significant (Table 50).

The percentage change between maximum and minimum value approached significance, again with a lower value in the high accident-rate curve. This near-significant result ($p = .06$, see Table 51) is attributable to the (non-significant) differences in maximum reaction time because the minimum reaction times before the curve did not differ (see additional Table 52).

Table 52. Differences in minimum reaction times between the high accident-rate curve A33 and the low accident-rate reference curve A36. Results for the t-test.

Pairs	*M*	*SD*	*r*	*p (r)*	*t*	*df*	*p*
A33	0.61	0.19					
A36	0.61	0.13	.46	.16	0.14	10	.90

Thus, the direction of the differences in reaction time is again not in line with the assumptions made in the hypotheses (irrespective of whether the difference was significant or not). However, the finding is in line with the general assumption of workload theories and the dual task paradigm: given the same amount of geometric demand, overall task demand is reduced if speed is reduced, which in turn – all other things being equal –results in reduced workload.

The significant lower reaction times in curve A33 which were attributed to the lower speed have another interesting implication for workload homeostasis theory. Although workload homeostasis theory does not directly assume a target level of workload analogous to the target level of risk in RHT (Wilde, 1994, 2001; Wilde et al., 1985), such a target level can be assumed because a homeostatic regulation implicitly requires a target level. This target level of workload would likely be situated at a medium level at which performance is highest (see Figure 8).

The significant differences in reaction time would only be in line with such an assumed target level of workload if speed was indeed unwillingly reduced by external influences like the speed camera in combination with the village. If speed were chosen freely, the difference in maximum reaction time would only be in line with this assumption if workload was still within a medium level and thus still within the target level despite significant differences. Other than the difference for curve A33 (Appendix A 4.5), no other differences within the group of high accident-rate curves were significant in terms of maximum reaction time. This fact allows the assumption that such a target level of workload might indeed exist and that the differences of curve A33 are indeed the result of the speed camera in combination with the village.

Because curve A24 was the same location as curve A33, but driven from the opposite direction, some of the assumptions made in the interpretation of the results regarding curve A33 can additionally be tested directly. A first assumption was that the lower speed in curve A33 can be attributed to the speed camera. This can be tested by comparing the two directions of location A24/A33 together with a comparison of the two directions of the reference location A21/A36. The results of this comparison indeed support this assumption (see additional Table 53 and Table 54): speed for the high accident-rate curve A33 was higher when driven from the direction without speed camera, whereas speed did not differ between the two directions of the reference curve.

The second assumption formulated above was that the lower speed in curve A33 caused significantly lower reaction times in comparison to the other high accident-rate curves as well as to the low accident-rate reference curve. This can again be tested by further analysing the values for reaction time, comparable to the comparisons for speed shown in Table 53 and Table 54. A similar pattern in the results can be interpreted as support for this second assumption.

Table 53. Differences between maximum, minimum, and percentage change in speed between the two directions of the high accident-rate location A24/A33. Results for the *t*-tests.

Parameter (Curve)	*M*	*SD*	*r*	*p* (*r*)	*t* (*df* = 8)	*p*
Max (A24)	79.66	7.35				
Max (A33)	67.48	5.47	.74	.02	7.38	.00
Min (A24)	67.16	6.58				
Min (A33)	63.56	4.87	.63	.07	2.09	.07
% Change (A24)	15.60	5.29				
% Change (A33)	5.69	4.08	.11	.79	4.69	.00

Table 54. Differences between maximum, minimum, and percentage change in speed between the two directions of the reference location A21/A36. Results for the *t*-tests.

Parameter (Curve)	*M*	*SD*	*r*	*p (r)*	*t* (*df* = 8)	*p*
Max (A21)	85.95	11.04				
Max (A36)	85.16	7.68	.63	.07	0.27	.79
Min (A21)	70.89	9.07				
Min (A36)	74.31	5.62	.33	.39	-1.15	.29
% Change (A21)	17.24	7.56				
% Change (A36)	12.48	5.56	.08	.84	1.58	.15

The results shown in Table 55 and Table 56 support this second assumption: for curve A33, in which speed was also lower, maximum reaction time and percentage change in reaction time were lower compared to its opposite direction. No differences were found between the two directions of the reference curve, which also showed no differences in speed. Regarding the pair-wise comparison with the low accident-rate reference curve A21, the high accident-rate curve A24 showed a lower maximum speed before the curve (Table 45) but no differences in minimum speed in the curve (Table 47) or in the percentage change in speed (Table 46). Thus, even for the opposite direction of curve A33, the results were again not in line with the hypotheses. Taken together, the behavioural differences found for the location A24/A33 do not explain the high amount of accidents that happened in this location.

Maximum speed before the curve and minimum speed in the curve were higher for curve A37 when compared to all other high accident-rate curves (Appendix A 4.5). These differences are probably attributable to the higher radius of curve A37 (see Appendix A 4.5): the radius for curve A37 was 198 m in comparison to radii between 128 m and 141 m for the other high accident-rate curves. Given these differences, the differences which were found between curve A37 and the low accident-rate reference curve A21 can also be explained by the radius, which was 135 m for the reference curve.

Therefore, although a higher maximum speed before the curve (Table 45) and a higher minimum speed in the curve (Table 47) were found for curve A37 in comparison to the low accident-rate reference curve A21, these results cannot be interpreted in favour of the hypotheses. This interpretation is supported by the significantly lower percentage change in speed for curve A37 (Table 46), the non-significant differences in longitudinal deceleration (Table 48), and the non-significant differences in maximum and percentage change in reaction times (Table 50 and Table 51).

Table 55. Differences between minimum, maximum, and percentage change in reaction time between the two directions of the high accident-rate location A24/A33. Results for the *t*-tests.

Parameter (Curve)	*M*	*SD*	*r*	*p (r)*	*t (df* = 8)	*p*
Max (A24)	0.67	0.14				
Max (A33)	0.67	0.19	.42	.26	0.10	.92
Min (A24)	2.12	1.01				
Min (A33)	1.03	0.47	.60	.09	4.02	.00
% Change (A24)	61.60	17.37				
% Change (A33)	30.95	12.12	.62	.08	6.68	.00

Table 56. Differences between minimum, maximum, and percentage change in reaction time between the two directions of the reference location A21/A36. Results for the *t*-tests.

Parameter (Curve)	*M*	*SD*	*r*	*p (r)*	*t (df* = 8)	*p*
Max (A21)	0.72	0.22				
Max (A36)	0.58	0.13	0.33	0.38	1.87	0.10
Min (A21)	2.12	1.18				
Min (A36)	1.41	1.00	-0.32	0.40	1.19	0.27
% Change (A21)	59.31	17.21				
% Change (A36)	46.69	22.10	-0.63	0.07	1.06	0.32

The characteristic difference of curve A50 compared to all other high accident-rate curves is its higher percentage change between maximum speed before the curve and minimum speed in the curve (Appendix A 4.5). This is clearly attributable to the speed limit sign at this location (see chapter 4.4.3.3). This interpretation is supported by the results shown in Table 45, Table 46, and Table 47: in comparison to the reference curve, curve A50 showed no differences in maximum speed before the curve, but significant differences in minimum speed in the curve and in the percentage change in speed. Furthermore, this higher change in speed was obviously planned well ahead and therefore cannot be attributed to drivers being surprised as this would have caused abrupt braking. This is shown by the non-significant differences in longitudinal deceleration, both in comparison to the reference curve (Table 48) and in comparison to the other high accident-rate curves (see Appendix A 4.5), unless they can be attributed to the peculiarities of the curves as discussed above.

In comparison to the reference curve, the high accident-rate curve A14 did not show any significant differences at all. Because curve A14 is not characterised by known peculiarities like the curves A33, A37 or A50, it is likely that these non-significant results represent what would have been found if the high accident-rate curve and the reference curve could have been perfectly matched in terms of their geometry and the geometry of the immediately preceding road section.

The results can be summed up as follows: while the majority of differences in behaviour were not significant at all, the few which were significant could be explained by peculiarities of the road sections. Thus, they could not explain differences in accident occurrence, nor did they support the assumptions formulated in the hypotheses. Given these results, some questions arose:

- Are the remaining non-significant differences attributable to weaknesses in the design of the study or are they representative?
- If the results were representative, why did accidents happen more often at the high accident-rate locations?
- Do the results question the validity of the model assumption?

Regarding the first question, there were some weaknesses in the study. Firstly, these concern the comparability of the high accident-rate curves and the reference curves and were already discussed above and in chapter 4.4.3.3. Secondly, they concern the sample size in terms of both the number of curves and – more importantly – the number of participants in the study. Of course, an increase in the number of participants increases the likelihood of differences becoming significant which were not significant with the smaller sample used here (Bortz, 2005). However, the descriptive statistics found for the small sample also allow the assumption that significant differences – if they were found with a larger sample – would similarly point towards different directions.

Thus, regardless of whether the non-significant results were representative or not, the implications were the same: the differences in the direction of the significant results would still not allow a statement regarding the effect of behaviour on accident occurrence between two curves with similar geometry. Given this finding, the other two questions posed above remain to be answered. Because of their importance with respect to the field study in general and the entire thesis, these are discussed in chapter 4.4.5 where the entire field study and its implications are discussed.

4.4.4.6 Results for Additional Research Question 3: Are There Differences in Gaze Between High and Low Accident-Rate Curves?

In line with the results shown in chapter 4.4.4.1 ('Effect of PDT on behaviour'), the idea of analysing differences in gaze behaviour between high and low accident-rate curves was not followed any further. This is because the PDT obviously had a levelling effect on potential differences in gaze behaviour between high and low accident-rate curves. The data were nevertheless analysed as done in the preceding comparisons with t-tests for paired samples but the results are only shown in Appendix A 4.6. As was expected, differences in gaze behaviour between high and low accident-rate curves were not significant. Because these non-significant results are due to the PDT and not due to a lack of differences in the curves, they have no implication for the original research questions and are therefore not further discussed.

4.4.4.7 Are Workload and Speed Related?

This chapter was conducted to find out if, and if so how, speed and workload are related. Background information on the expected finding was given in chapter 4.4.2, where the hypotheses for the field study were developed. In order to examine this question, speed and reaction-time values were first averaged across all participants. Thus, each curve could be described with a single value in each of the parameters used for the subsequent analyses.

The resulting values were firstly used for preliminary tests of how geometry and the two behavioural variables of speed and reaction time in the PDT are related. This was done by performing linear regression analyses. In these analyses geometry represented the independent variable and either speed or reaction time represented the dependent variable. Geometry was represented by three parameters:

- the minimum radius in the section (min. R);
- the curvature change rate of the curve with the minimum radius (CCRs); and
- the curvature change rate of the entire section including approximately 150 m before curve beginning (CCR).

These parameters were calculated for all 21 sections which were driven with PDT. These 21 sections included all high accident-rate curves together with the reference curves used in the preceding chapters. Most sections were single curves which were preceded by longer straight road sections. Accordingly, CCR was not calculated for the sections preceding the curves, but is indirectly included in the CCR for the entire section. In addition to the single curves, two longer sections with several curves and consequently high CCR were included: section A16 with a CCR of 389.2 and section

A43 with a CCR of 387.9. For those two sections the curve with the smallest radius was used for the calculation of the CCRs and minR values. Regarding the behavioural variables speed and reaction time, two parameters were used for each variable:

- minimum speed and maximum reaction time as singular parameters, and
- the percentage change in speed and reaction time between maximum and minimum value across the entire section as a combined parameter which also takes into account the characteristics of the preceding road section.

Before statistically analysing the values, all dependent and independent variables were tested with the K-S test, which indicated no deviation from the assumption of normal distribution. Because it is unknown whether and how speed and reaction time are related in this data, it was decided to calculate separate regression analyses for each variable; otherwise problems of multicollinearity would have been likely (Backhaus et al., 2006; Brosius, 2008). This is definitely the case for the different parameters of geometry and the different parameters of the behavioural variables. Of course, this resulted in a multitude of regression analyses for which the most important results are shown in Table 57 and Table 58.

Table 57. Results of several linear regression analyses of different geometric parameters on minimum speed and percentage change in speed.

Parameter	*B*	*β*	*t*	*p*	R^2	*F*(1, 19)	*p*
			Minimum speed				
min. R	0.03	.78	5.39	.00	.61	29.05	.00
CCRs	-0.04	-.78	-5.41	.00	.61	29.23	.00
CCR	-0.09	-.84	-6.61	.00	.70	43.72	.00
			Percentage change in speed				
min. R	-0.01	-.59	-3.17	.01	.35	10.07	.01
CCRs	0.02	.66	3.84	.00	.44	14.73	.00
CCR	0.02	.32	1.48	.16	.10	2.19	.16

Summing up Table 57 and Table 58, geometry did indeed predict both speed and reaction time rather well. The quality of the regression was best for minimum speed. This is not surprising because a large portion of these results could be explained by the physics of driving (see chapter 4.3.3.12). The results further indicate that the changes in speed and reaction time (RT) were related.

Table 58. Results of several linear regression analyses of different geometric parameters (minimum Radius, CCR$_S$ and CCR) on maximum reaction time and percentage change in reaction time.

Parameter	*B*	*β*	*t*	*p*	R^2	*F*(1, 19)	*p*
			Maximum reaction time				
min. R	0.00	-.61	-3.37	.00	.37	11.37	.00
CCRs	0.00	.45	2.20	.04	.20	4.84	.04
CCR	0.00	.26	1.17	.26	.07	1.37	.26
			Percentage change in reaction time				
min. R	-0.03	-.65	-3.74	.00	.42	13.98	.00
CCRs	0.03	.46	2.27	.04	21	5.13	.04
CCR	0.05	.33	1.54	.14	.11	2.37	.14

This potential relationship was tested both for the single behavioural indicators minimum speed and maximum RT and also for the combined behavioural indicators percentage change in speed and percentage change in RT. Again, linear regression analyses were performed. Reaction time was chosen as the dependent variable and speed as the independent variable. From a statistical point of view, the combination could also have been vice versa. The results for the two regression analyses are shown in Table 59; the results for the percentage change are also visualised in Figure 49. The results are summerised and discussed in the next chapter.

Table 59. Results of two linear regression analyses (minimum speed on maximum RT and percentage change in speed on percentage change in RT).

B	*β*	*t*	*p*	R^2	*F*(1, 19)	*p*
			Minimum speed on maximum RT			
-0.03	-.47	-2.33	.03	.22	5.42	.03
			% change in speed on % change in RT			
1.78	.79	5.68	.00	.63	32.28	.00

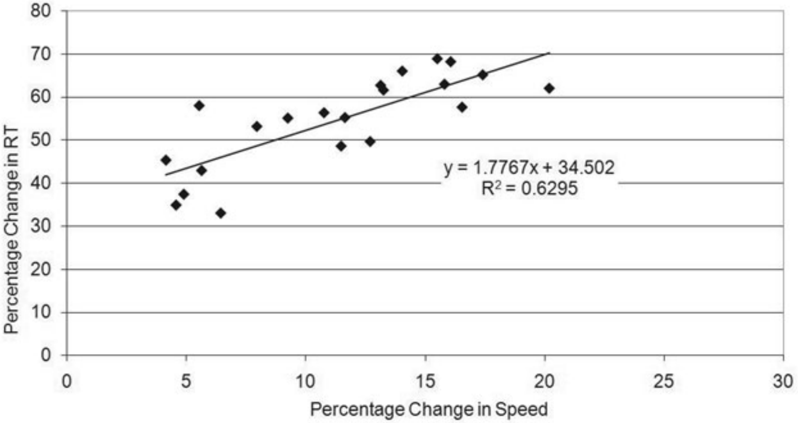


Figure 49. Relationship between percentage change in speed and percentage change in reaction time, measured for several curves.

4.4.4.8 Summary and Discussion of Results for the Relationship Between Speed and Reaction Time

The findings are interesting with respect to the theoretical part of this work. Perhaps most important in this sense is that both speed and reaction time were influenced by road geometry. For speed alone, this finding is not very spectacular because the effect can be explained by the laws of physics (see chapter 4.3.3.12). These also explain why the quality of the regressions was better for speed than for reaction time. However, this is only valid when the parameter minimum speed is used, which is to a certain extent self-explanatory. On the other hand, the regression of speed on reaction time was better for the parameter percentage change than for the parameters minimum speed and maximum reaction time. This could mean that the percentage change is a more valid predictor of workload than the singular parameters minimum and maximum, which – at least for speed – tend more to reflect road geometry alone.

Nevertheless, the fact that objective road geometry had such a significant influence on both speed and reaction time explains why differences in behaviour between two curves which are similar in terms of their geometry but which differ in terms of their accident rate are hard to find (see preceding chapters). The fact that reaction time is influenced by road geometry, despite speed being reduced, has further important implications with respect to the theories and is in itself a highly important finding.

Firstly, this shows that both the regulation of speed and the reaction to the PDT share the same resources (Wickens, 1984, 2008). From a methodological point of view, this means that the PDT used in the driving experiments was indeed a valid predictor of this kind of workload that results from demand imposed on the driver by road geometry. At the same time, the question arises as to why the PDT is needed, if measuring the reduction in speed alone allowed similar statements. This question was already answered in chapter 4.4.4.3 where it could be shown that only the additional demand of the PDT resulted in a clear distinction in speed between different locations. This means that speed regulation without additional PDT might not have been sensitive enough to changes in perceived demand.

Secondly, the fact that reaction time was influenced by road geometry, despite speed being reduced, is interesting with respect to the predictions made by workload homeostasis theory (see chapter 2.3.5). According to this theory, drivers could have been expected to decrease speed with increasing demand in order to maintain their target level of workload (see also preceding chapter). If this was valid, the regression of geometry on reaction time (Table 58) should not have been significant. The fact that it was shows that workload homeostasis was not achieved in curves. A failure in such homeostatic control could be interpreted as a failure to keep workload at an optimum level. This increase in reaction time might even indicate that a certain workload red-line was exceeded. In this case the task capability interface model of Fuller (2005, see Figure 10) predicts an increase in the likelihood of accidents. In fact, this is what is found when curve geometry and accident occurrence are related (see chapter 2.3.9.3).

Thirdly, the fact that the percentage change in reaction time increased with the percentage change in speed supports the argument above: a high change in speed is usually associated with a higher likelihood of an accident (L. Aarts & Schagen, 2006). This and the considerations in the preceding paragraph indicate that accident occurrence could be predicted by workload.

However, all these findings with respect to accident occurrence only apply on an aggregated level. As was shown in the preceding chapters, a distinction between high and low accident-rate curves was not possible once geometry was accounted for. Here, other aspects could be more important and require further investigation. However, the aggregated results concerning workload and accident occurrence did not differ much from results reported when road geometry alone was used. Thus, the results found here are of an astonishingly high quality, given that not geometry but behavioural variables, which differ between drivers, were used.

It might even be possible for these aggregated results to be further improved. This is because of certain methodological shortcomings that could not be avoided for the data at hand. These shortcomings regard the sample size in terms of both curves and participants.

With respect to these shortcomings, the quality of the aggregated data for each curve could be improved with a larger sample size. This is because, for the analyses above, the data had to be aggregated for all participants for whom valid data was available for this particular curve.

Another aspect of data quality concerns the calculation of the values characterising the section preceding the curves. In order to have comparable data for all curves, the section used to calculate these values was restricted to approximately 150 m before the curves. This value was chosen because it was the minimum distance for some curves at which the PDT started. Although this distance is rather short, it ensured at the same time that the direct influence of the curve under consideration was used and not the effect of some other element preceding the curve.

4.4.5 Summary and Discussion of the Results for the Field Study

The field study was mainly conducted to test the final part of the model, that is, the effect of behaviour on accidents. The research paradigm used to analyse this effect was a comparison between curves which were similar in terms of their geometry but which differed in terms of their accident rate. Because behaviour in the curve with no accidents was deemed as appropriate behaviour (see the driver and driving behaviour model for rural roads, Figure 14), differences in behaviour between high and low accident-rate curves could be interpreted as supporting the model assumptions. In addition, such findings have important implications for the future assessment of road safety. For example, it might be possible to infer accident likelihood in curves based on behaviour. Because geometry was comparable in both types of curves, this approach exceeded current approaches that mainly use geometry and parameters of traffic density to predict accident occurrence (overview of this and other approaches in Reurings et al., 2005).

It was assumed that high accident-rate curves, when compared to geometrically similar curves with no accidents, were characterised by higher speeds, higher lateral acceleration and higher longitudinal deceleration. Additionally, tests were used to determine whether speed before the curve differed between high and low accident-rate curves. In this study, behaviour was not restricted to driving behaviour but also took into account workload and gaze behaviour. Workload was assessed with a peripheral detection task (PDT); gaze behaviour was recorded with a contact-free eye-tracker. In accordance with the task capability interface model (Fuller, 2005), it was assumed that workload and the percentage change of workload between maximum and minimum were higher in high accident-rate curves. However, it was also assumed that some kind of trade-off existed between speed and the performance in the peripheral detection task.

Therefore, the effect of the PDT on driver behaviour had to be analysed first in order to understand and correctly interpret the results with respect to the hypotheses. This analysis was not restricted to single curves (chapter 4.4.4.1) but took also into account different geometric characteristics of several curves (chapter 4.4.4.7). The latter served as input for additional testing of assumptions in relation to workload homeostatic theories of driving behaviour. Similar to the effect of the PDT on speed, an effect of the PDT on gaze behaviour was assessed. This was conducted to determine the extent of the PDT influence on potential differences in gaze behaviour between high and low accident-rate curves.

Concerning the effect of the PDT on driving behaviour, it was found that the PDT in general resulted in a decrease in speed. In addition, the PDT tended to increase differences in speed between curves of different geometry. This non-significant interaction effect resulted in comparable values for reaction time when the curves were driven in the PDT condition. This result at first sight seemed to support a workload homeostatic way of speed regulation. However, this interpretation was not supported when reaction time values were compared between the straight road section before the curve and the curve itself. If workload homeostasis were applicable to curve driving, these values should not have differed. The fact that they did (clearly visible and thus not tested statistically) was statistically supported when several curves of different geometric demand were included in an analysis of the trade-off between speed and reaction time. Here, it was indeed found that reaction times increased with a decrease in speed which in turn was caused by a decrease in radius.

Summing up these findings, some support for a workload homeostatic control of speed could be found. However, this homeostatic control was not achieved between elements that differ fundamentally in terms of their geometry such as straight road sections versus curves. In fact, the higher the demand of a curve reflected by lower average speed across all drivers, the less likely that workload can be kept at an acceptable level. This finding is in accordance with the well-known relationship between curve geometry and accident occurrence (see chapter 2.3.9.3). In fact it exceeds this known relationship because it is based on actual behaviour and thus could be more relevant to the prediction of accident occurrence than solely using geometry. This interpretation is supported by the findings of the simulator study, where it was found that several other design elements besides road geometry determine driving behaviour.

Regarding the effect of the PDT on gaze behaviour, it was found that it had a significant influence on all gaze parameters. Both the standard deviation of the gaze data and the scan paths differed significantly between the condition with and without PDT. The standard deviation of the gaze data (SD-gaze) was additionally analysed due to its higher eta squared in the ANOVA. By performing binary logistic regression analysis, it was found that SD-gaze could reliably distinguish between the condition with and without PDT. This finding was independent of the environmental differences present in different curves. Despite not being directly related to the topic of this thesis, this result is highly important and encouraging with respect to the development of a future in-vehicle device that could be designed as IVIS or ADAS to detect driver inattention or distraction.

With respect to the PDT-influence, gaze variables other than the scan path and SD-gaze were more difficult to interpret. It was assumed that the PDT had interacted with the area used for the calculation of some of these variables, the data quality in general, and the spatial and temporal limits used to define fixations. Even more important with respect to the differences between high and low accident-rate curves was the finding that none of the gaze variables could distinguish between the curves used to analyse the influence of the PDT. This means that applying the PDT had such a vast influence on the drivers' gaze behaviour that apparent differences between curves (see Figure 38) were levelled out. While such a finding is important for certain IVIS and ADAS applications (see above), it also meant that gaze data could not be used to distinguish between high and low accident-rate curves, at least not when the PDT was being attended to. Therefore, it was decided not to further analyse gaze data for differences between high and low accident-rate curves.

The central focus of the driving experiments was to identify differences in behaviour which could be used to explain and predict accident occurrence according to the relationships assumed in the driver and driving behaviour model for rural roads. The data analysis showed very few significant differences in behaviour between high and low accident-rate curves. Furthermore, the few differences which were found were either not in accordance with the model or could be explained by external influences such as minor differences in geometry or different characteristics in the approach section. It had to be concluded that the relationship between behaviour and accidents proposed by the driver and driving behaviour model for rural roads could not be confirmed once geometry was accounted for. In chapter 4.4.4.5, questions were posed as to how these results could be explained and whether they threatened the validity of the model. The following paragraph answers these questions.

The approach used in this thesis was based on specific assumptions concerning accident occurrence. One such implicit assumption was that the probability of an accident increases at high accident-rate locations because of an overall change in average behaviour towards more dangerous behaviour. This idea is visualised in Figure 50 in which the 'behaviour red-line' indicates an assumed limit above which accidents happen. In Figure 50 it is assumed that because of this general shift in average behaviour, a further shift in individual behaviour towards more dangerous behaviour increases the accident risk at high accident-rate locations. This is not the case in low accident-rate locations, although the level of deviation in individual and average behaviour is the same.

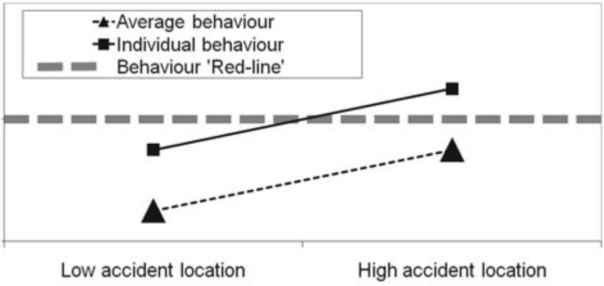


Figure 50. A general idea behind accident occurrence.

Relationships as depicted in Figure 50 are quite likely to be found if the high accident-rate location is designed contrary to quality standards defined by road engineers such as those found in the RAS-L (FGSV, 1995) or by general principles of good design stated by self explaining road principles (Theeuwes, 2000; Theeuwes & Godthelp, 1995) or more generally by Norman (1998). Given the relatively high standard of road safety in Europe, very few real problem locations of this kind ('accident black spots') can be expected. With today's knowledge, their existence would rather point towards severe neglect or shortage of resources on the part of the responsible road authority. However, accidents can also occur without a shift of average behaviour, which is shown in Figure 51.

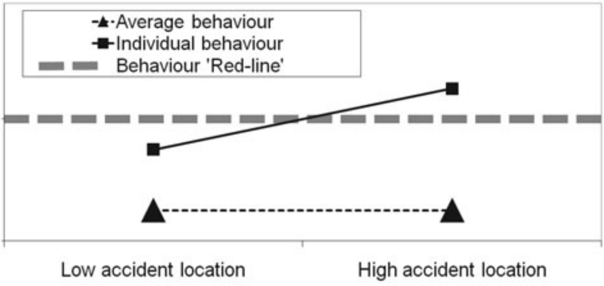


Figure 51. Alternative idea behind accident occurrence.

In Figure 51, the high accident-rate location has characteristics which induce very inappropriate behaviour in only very few drivers (termed 'individual behaviour') but do not change average behaviour. In order for such behaviour to occur, the high accident-rate location has to permit such inappropriate behaviour, that is, it has to be characterised by a high degree of freedom. This might be the case at locations which are only regulated by formal speed limits: most drivers would adhere to these limits, but the very few who did not realise that there was a speed limit at all might cause an accident.

This second idea would also explain why accidents could occur despite the non-significant differences in the sample, regardless of whether these are representative or not (see chapter 4.4.4.5). It would also explain why no differences in speed were found in similar studies conducted in the laboratory with few participants (Shinar, 1977).

It must be noted that this alternative idea of accident occurrence is difficult to examine in experimental studies. Of course, instead of accidents themselves, driving errors or traffic conflicts could be used as proxy variable for accidents. However, even in this case, a very large sample of drivers would be needed to collect enough errors or conflicts for a meaningful statistical analysis between two different locations. By using assumed conversion factors of 1/10,000 (Gstalter, 1983, cited in Reichart, 2001), or 2.4/51,100 (Dingus, Hetrick, & Mollenhauer, 1999) between an accident and a traffic conflict or a 'Driver error; Hazard present' (Dingus et al., 1999), a sample size of between 23 to 50 participants would be needed for the locations used in the field study to get just a single conflict. However, this single conflict would not be enough for a statistical analysis.

Therefore, if accidents in a high accident-rate location could be explained by the alternative idea depicted in Figure 51, in-depth accident analyses on the spot would be the method of choice. This could also include analyses of data stored in automatic-data recording devices integrated into modern cars. Such in-depth studies should put additional emphasis on psychological variables. For the example introduced above this means that questions should be asked as to why a speed limit sign was not perceived. The natural next question would then be whether the sign should or could be replaced by a more self-explanatory speed limit design feature as was shown in the simulator study. Concerning the implication for the model's validity, the results neither supported nor falsified the model assumptions with respect to accident occurrence. Until additional studies are conducted as described above, the model assumptions should be kept.

5 Empirical Validation: Summary and Conclusions

Rural roads are the most dangerous road category when the number of fatal accidents is used as the criterion. In order to improve safety on rural roads, a driver and driving behaviour model was developed in a first step. In this model, a relationship between perception, behaviour and accidents is proposed that allows the formulation and testing of hypotheses. The conclusions derived from the testing of these hypotheses were expected to increase insight into how behaviour is determined on rural roads and how rural road safety can be increased. The empirical validation of this model was the topic of this thesis, for which three (quasi-) experimental studies were conducted:

- a laboratory study;
- a simulator study; and
- a field study.

Each of these empirical steps served to test a specific model assumption.

The laboratory study was conducted as a preparatory study with respect to the influence of the perceived road situation ahead on behaviour. One aspect of this study was to determine whether the existing Road Environment Construct List (RECL) (Steyvers, 1993, 1998; Steyvers et al., 1994) could be used to assess this part of the model which was termed 'perceived road situation ahead'. In order to answer this question, a German version of the RECL and its factor structure had to be validated. The material used in the laboratory study were 21 pictures of a large variety of two-lane rural roads. These 21 pictures were rated by 46 participants with the translated RECL items and an additional item asking for the preferred speed on the roads shown on the pictures.

The analysis of the factor structure revealed that two concurrent solutions would be permissible on statistical grounds: a two-factor solution and a three-factor solution. However, even the three-factor solution showed some minor but important deviations from the original RECL solution (Weller, Schlag et al., 2008). This was even more the case for the two-factor solution discussed in this thesis. In addition to the data collected in the laboratory study, the data collected in the simulator study was used to analyse the factor structure. The resulting structure again deviated from the structure found in the laboratory study and the original RECL structure. Therefore, instead of using the original RECL factors, it was decided to select a few marker items based on the combined results of the factor analyses and existing theories of driver and driving behaviour. The two marker items ‘monotonous’ and ‘dangerous’ allowed a highly significant prediction of rated speed in the laboratory task.

In an additional step, these two marker items were used to assess the influence of individual factors on perception and behaviour. Up to this point, individual variables were only implicitly integrated into the model and thus this analysis served to establish whether their influence in the model had to be stressed further. As a result of preliminary analysis, the demographic variable age was selected to represent individual variables. How this individual variable influenced perception and behaviour was prototypically tested by applying the methods of structural equation models to the data collected in the laboratory. It was found that a direct path of age towards behaviour in addition to an indirect path via perception as proposed in the model is likely. However, in view of the nature of the data (laboratory task, ratings of behaviour instead of real behaviour), it seemed advisable not to change the model at this point without further support of the data collected in the simulator study.

The simulator study constituted an extension of the laboratory study. While it again served to assess the influence of the perceived road situation ahead on behaviour, actual behaviour could be recorded instead of merely being rated. The experimental setting also allowed hypotheses to be formulated and tested. These hypotheses concerned the role of single cues in curves and of different environments of straight road sections and the affordances provided by both. The study was conducted in the simulator of the Fraunhofer IVI in Dresden with 50 participants who drove a course of nine kilometres in two directions.

In order to ensure that the data collected in the simulator was a valid and reliable predictor of real behaviour in the field, several additional steps were conducted prior to analysing the data with respect to the model. Amongst these steps were

- the discussion of the role of simulator sickness;
- the discussion of the issue of insufficient familiarity and adaptation of behaviour in the simulator;
- the assessment of the potential influence of element order in the simulated course; and
- the preliminary assessment of the ecological validity of the simulator used in this study.

The findings were used to select appropriate parameters for the analysis of the data. They can also be used to develop a prototypical check-list for future simulator studies. This seems necessary given the somewhat scant way in which the issues enumerated above are sometimes reported in publications dealing with simulator studies.

Regarding the validation of the model, the results of the simulator study were in accordance with some, but not all, of the assumptions formulated in the hypotheses. They were in line with the assumptions concerning the effect of single cues and various environments on behaviour. Environmental characteristics influence behaviour in addition to influences of geometry alone. This has direct relevance for traffic safety as will be discussed below.

The assumption that behaviour is the result of a conscious evaluation of the perceived road situation ahead and the subsequent expected appropriate behaviour for this situation was only found for curves, not for straight road sections. For the latter, the ratings used as proxy for perception and expectation were even misleading and had no relation to measured speed. Behaviour on straight road sections could instead be explained by the optic flow as assumed in Gibson's direct approach to perception (Gibson, 1986). Thus, two different perceptual processes were found to explain behaviour: a direct effect of the environment and an indirect effect that requires conscious processes. How can this finding be incorporated into the model?

Based on the results, it could be assumed that these two processes are unique to either curves or straight road sections. However, this distinction is not regarded as appropriate. This is because the optic flow is present regardless of road geometry and, similarly, expectations can also develop on straight road sections, although they were found to be misleading in the present study. However, the findings can be integrated into the model by assuming an attentional monitor that constantly checks whether the situation ahead differs from the current situation. Such an assumption is common in psychological models such as the GEMS model by Reason (Figure 2) or the zero-risk model proposed by Näätänen & Summala (see chapter 2.3.4).

A need for conscious information processing only arises if the situation ahead differs from the current situation. In this case, open-loop control of behaviour permits an estimation of appropriate behaviour for the situation ahead. Open-loop adaptation to behaviour is made by comparing current and expected behaviour. If the situation ahead does not differ from the current situation, behaviour is regulated in a closed-loop way by the optic flow and perceptual invariants. Forcing participants to switch to open-loop control in such closed-loop situations leads to a mismatch between ratings and behaviour as was found for the straight road sections.

The comparison of current situation and situation ahead can be done by the three dimensions named in the original model:

- the objective road geometry and situation;
- affordances and cues; and
- individual knowledge, experience and mental models.

It could be that these dimensions are translated into ratings of risk as is assumed in the zero-risk model of Näätänen & Summala (1976). However, at this stage there is no need for this step as an activation of open-loop control can even be done by the basic distinction between curves and straight road sections. The results indicate that such specific rating is only done after it is decided that current situation and future situation do not match. Once open-loop control is activated, subjective risk and subjective demand are the relevant variables that determine expected appropriate speed for the situation ahead.

It was found that demographic variables only influenced measured behaviour and not perception or expectations which were collected with ratings. This finding was independent of whether curves or straight road sections were analysed. It indicates that demographic variables influence behaviour in a rather late stage of the model. Mathematically, demographic variables can be viewed as calibration factors that are used to transform open-loop and closed-loop perceptual information into behaviour. Although the results of the simulator study do not support a mediation model, they are consistent with the preference for the partial mediation model instead of the complete mediation model found in the laboratory study. In terms of the validity of the results for this specific question, the simulator data are regarded as more relevant to the model than the results found for the laboratory data. This is due to the fact that in the latter case the exploratory analyses of differences between driver groups were restricted to a selected subset of roads and did not include measured driving behaviour.

The combined findings of the simulator study and of the laboratory study led to modifications in the original driver and driving behaviour model for rural roads. This modified model is depicted in Figure 52.

What do the findings so far mean in combination with the modifications of the model for traffic safety? Most important is the fact that behaviour can be influenced without changing road geometry. This applies to both open-loop and closed-loop control. For closed-loop control, changing the optic flow and the values of perceptual invariants results in respective changes in behaviour. For open-loop control, an effect can be achieved by influencing perceived risk associated with the situation ahead. For example, it was found that a reduction in sight distance could reduce speed in curves by a similar amount to when formal curve warning signs are present. Both were found to be rated similarly with respect to rated risk. By deliberately increasing perceived risk, speed can be reduced and safety can be increased.

Finally, a field study was conducted to test the last part of the model, that is, the effect of behaviour on accidents. The research paradigm used to analyse this effect was to compare pairs of curves which were similar in terms of their geometry but which differed in terms of their accident rate. Behaviour in the curve with a lower accident-rate was equated with ‘appropriate behaviour’ as named in the model. In addition to speed referenced to the road location, behaviour was also assessed with a contact-free eye-tracker and a peripheral detection task (PDT). The inverse of the reaction time assessed with the PDT was used as proxy variable for workload. The data were collected for 16 participants who drove a test-route in the outbound and inbound direction, amounting to a total of 80 kilometres. Four high accident-rate locations were situated within this course.

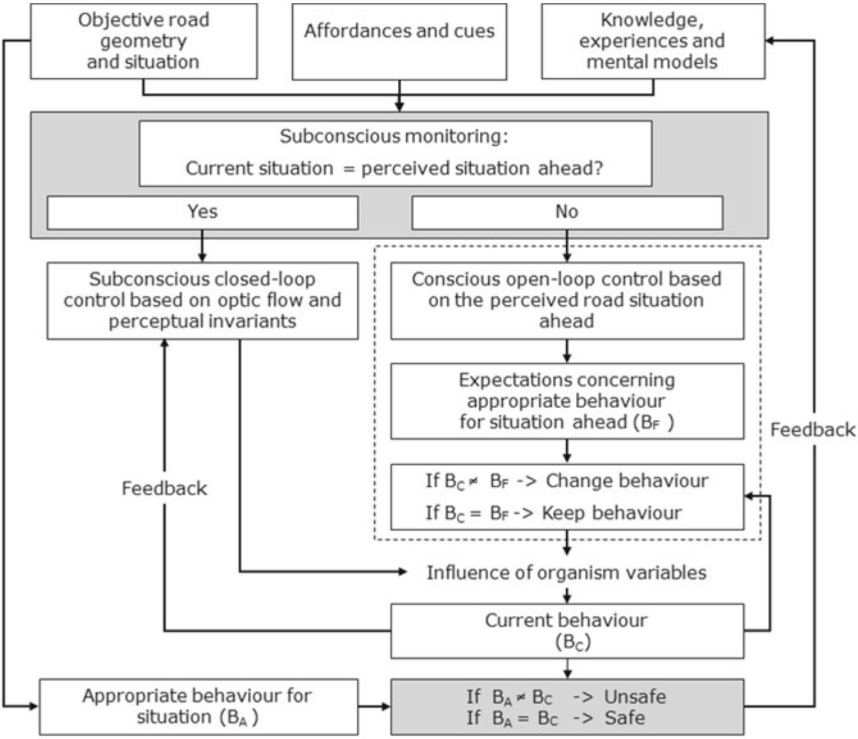


Figure 52. Modified driver and driving behaviour model for rural roads.

Similar to the simulator study, several issues had to be addressed as prerequisite for a meaningful interpretation of the results. Firstly, issues in relation to the familiarization with the PDT were analysed. This analysis was followed by an assessment of the effect of the PDT on 'normal' behaviour without the PDT. Finally, the quality of the gaze data was determined.

The findings regarding the effect of the PDT on 'normal' behaviour first of all indicated that the specific PDT used in this study was indeed a valid method to assess workload in driving (Wickens, 1984, 2008). However, it was also found that the PDT had such a vast influence on gaze behaviour that differences between curves were completely levelled by the PDT. Therefore, differences in gaze behaviour between high and low accident-rate curves were not further analysed.

Although the strong effect of the PDT on gaze behaviour ruled out its analysis with regard to the hypotheses, it did allow novel approaches to be pursued. This was done by performing binary logistic regressions with several parameters of gaze behaviour. It was shown that the standard deviation of the gaze data reliably distinguished between the condition with and without PDT. This finding is regarded as highly important for application in future driver information or driver assistance systems (IVIS & ADAS) that allow detection of driver inattention or distraction.

The results did not support the hypotheses regarding differences in behaviour between high and low accident-rate curves with similar geometry. The reason for this finding is seen in a proposed alternative explanation of accident occurrence: while the field study was conducted with the assumption that average behaviour shifts towards more dangerous behaviour in high accident-rate curves (Figure 50), the alternative explanation assumes that average behaviour does not change (Figure 51). What does occur, however, is a dramatic change of behaviour in very few individuals. According to the alternative explanation, it is this very inappropriate behaviour of the very few that results in accidents. Detecting such rare events in behaviour in a field study is difficult and would have required a very large sample of participants. Although the model could not be validated with respect to accident occurrence, the findings did not falsify it either. It is therefore proposed that its validity remains assumed until future in-depth studies are conducted.

This decision is indirectly supported by the analysis of the relationship between workload, speed and geometric demand at an aggregated level. Here it was found that both speed decreased and reaction time increased with geometric demand. Because accident likelihood also increases with geometric demand, this finding means that a similar relationship also existed for behaviour and accident occurrence. In fact, for the field study data it was found that geometric demand predicted both speed and reaction-time parameters similarly well. The fact that the percentage change in speed between maximum speed before the curve and minimum speed in the curve could predict the percentage change in workload means that predictions based on workload homeostasis theory were falsified. Thus, this theory cannot be applied to curve driving. Given the influence of homeostatic theories in driving (Fuller, 2005; Gstalter & Fastenmeier, 1995; Wilde, 1988, 2001), it is suggested that additional customised studies for the testing of these theories in curve driving be conducted.

Summing up the empirical validation of the original model (Figure 14), an overall effect of perceptual processes could be shown by the studies conducted in the laboratory and in the simulator. Unlike the original model, it was found that a distinction must be made between two perceptual processes (Figure 52):

- a closed-loop perceptual control of behaviour that is based on perceptual invariants and the optic flow, and
- an open-loop control based on conscious evaluation of the road situation ahead.

Behavioural differences resulting from these processes were found but could not be directly related to accident occurrence in the field study. However, analysis of the effect of geometric demand on behaviour at least allowed indirect support of the model assumptions regarding accident occurrence. For future validation of these assumptions it is proposed that experimental data collected in field studies be directly compared to data collected from in-depth accident analyses. The latter must be supported by data from automatic data-recording devices which are possibly available in today's modern cars.

References

Aarts, H., & Dijksterhuis, A. (2000). Habits as knowledge structures: Automaticity in goal-directed behavior. *Journal of Personality and Social Psychology, 78*(1), 53–63.

Aarts, L., & Schagen, I. v. (2006). Driving speed and the risk of road crashes: A review. *Accident Analysis & Prevention, 38*(2), 215–224.

Allport, D. A., Antonis, B., & Reynolds, P. (1972). On the division of attention: A disproof of the single channel hypothesis. *Quarterly Journal of Experimental Psychology, 24*(2), 225–235.

Amelang, M., Bartussek, D., Stemmler, G., & Hagemann, D. (2006). *Differentielle Psychologie und Persönlichkeitsforschung [Differential psychology and personality research]* (6th ed.). Stuttgart: Kohlhammer.

American Association of State Highway and Transportation Officials (AASHTO). (2001). *A policy on geometric design of highways and streets* (4th ed.). Washington, DC: AASHTO.

Anderson, K. J. (1990). Arousal and the inverted-U hypothesis: A critique of Neiss's 'Reconceptualizing arousal.' *Psychological Bulletin, 107*(1), 96–100.

Apter, M. J. (1984). Reversal theory and personality: A review. *Journal of Research in Personality, 18*(3), 265–288.

Arbuckle, J. L. (1997). *Amos user's guide version 3.6*. Chicago: Smallwaters.

Arthur, W., Strong, M. H., & Williamson, J. (1994). Validation of a visual attention test as a predictor of driving accident involvement. *Journal of Occupational and Organizational Psychology, 67*(2), 173–182.

Assum, T., Bjørnskau, T., Fosser, S., & Sagberg, F. (1999). Risk compensation - the case of road lighting. *Accident Analysis & Prevention, 31*(5), 545–553.

Avolio, B. J., Kroeck, K. G., & Panek, P. E. (1985). Individual differences in information-processing ability as a predictor of motor vehicle accidents. *Human Factors, 27*(5), 577–587.

Backhaus, K., Erichson, B., Plinke, W., & Weiber, R. (2006). *Multivariate Analysemethoden - Eine anwendungsorientierte Einführung [Multivariate analysis methods - an application-oriented introduction]*. Berlin: Springer.

Backs, R. W., Lenneman, J. K., Wetzel, J. M., & Green, P. (2003). Cardiac measures of driver workload during simulated driving with and without visual occlusion. *Human Factors, 45*(4), 525–538.

Ball, K., & Owsley, C. (1991). Identifying correlates of accident involvement for the older driver. *Human factors, 33*(5), 583–595.

Ball, K., Owsley, C., Sloane, M. E., Roenker, D. L., & Bruni, J. R. (1993). Visual attention problems as a predictor of vehicle crashes in older drivers. *Investigative Ophthalmology & Visual Science, 34*(11), 3110–3123.

Barry, R. J., Clarke, A. R., & McCarthy, R. (2005). Arousal and activation in a continuous performance task: An exploration of state effects in normal children. *Journal of Psychophysiology, 19*(2), 91–99.

Baumann, M., & Krems, J. F. (2007). Situation awareness and driving: A cognitive model. In C. Cacciabue (Ed.), *Modelling driver behaviour in automotive environments* (pp. 253–265). London: Springer.

Beauducel, A., Brocke, B., & Leue, A. (2006). Energetical bases of extraversion: Effort, arousal, EEG, and performance. *International Journal of Psychophysiology, 62*(2), 212–223.

Becher, T., Baier, M. M., Steinauer, B., Scheuchenpflug, R., & Krüger, H.-P. (2006). Berücksichtigung psychologischer Aspekte beim Entwurf von Landstraßen - Grundlagenstudie. [Psychological aspects in the design of rural roads - Literature review]. In Bundesanstalt für Straßenwesen (Ed.), *Berichte der Bundesanstalt für Straßenwesen. Heft V148*. Bremerhaven: Wirtschaftsverlag NW.

Bella, F. (2008). Driving simulator for speed research on two-lane rural roads. *Accident Analysis & Prevention, 40*(3), 1078–1087.

Berlyne, D. E. (1960). *Conflict, arousal, and curiosity*. New York: McGraw-Hill.

Berlyne, D. E. (1970). Novelty, complexity, and hedonic value. *Perception & Psychophysics, 8*(5-A), 279–286.

Berlyne, D. E. (1972). Ends and means of experimental aesthetics. *Canadian Journal of Psychology/Revue Canadienne de Psychologie, 26*(4), 303–325.

Betsch, T. (2005). Wie beeinflussen Routinen das Entscheidungsverhalten? [How do routines influence decision making?]. *Psychologische Rundschau, 56*(4), 261–270.

Borg, I., & Staufenbiel, T. (2007). *Lehrbuch Theorien und Methoden der Skalierung [Textbook theories and methods of scaling]* (4th ed.). Bern: Huber.

Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaftler [Statistics for the human and social sciences]* (6th ed.). Heidelberg: Springer.

Bortz, J., & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler [Research methods and evaluation for the human and social sciences]* (4th ed.). Heidelberg: Springer.

Bouchard, S., Robillard, G., & Renaud, P. (2007). *Revising the factor structure of the Simulator Sickness Questionnaire. Oral presentation at the 12th Annual CyberTherapy 2007 Conference: Transforming Healthcare Through Technology. June, 2007. Washington, D.C.* Retrieved September 26, 2008, from http://www.interactivemediainstitute.com/PPT/Cybersickness/bouchard.pdf

Boucsein, W. (1988). *Elektrodermale Aktivität. Grundlagen, Methoden und Anwendungen [Electrodermal activity. Basics, methods and application]*. Berlin: Springer.

Boucsein, W. (1992). *Electrodermal activity*. New York: Plenum Press.

Boucsein, W. (1995). Die elektrodermale Aktivität als Emotionsindikator [Electrodermal activity as indicator of emotion]. In G. Debus, G. Erdmann & K. W. Kallus (Eds.), *Biopsychologie von Streß und emotionalen Reaktionen [Biopsychology of stress and emotional reaction]* (pp. 143–161). Göttingen: Hogrefe.

Boucsein, W. (2001). Physiologische Grundlagen und Meßmethoden der dermalen Aktivität [Physiological basics and measuring methods of dermal activity]. In F. Rösler (Ed.), *Enzyklopädie der Psychologie. Themenbereich C: Theorie und Forschung; Serie 1: Biologische Psychologie; Bd. 4: Grundlagen und Methoden der Psychophysiologie [Basics and methods of psychophysiology]* (pp. 551–623). Göttingen: Hogrefe.

Boyce, T. E., & Geller, E. S. (2002). An instrumentes vehicle assessment of problem behaviour and driving style: Do younger males really take more risks? *Accident Analysis & Prevention, 34*(1), 51–64.

Brannolte, U., Barth, H.-B., Schwarzmann, R., Junkers, W., Liu, Y., Sigthorsson, H., et al. (1993). *Sicherheitsbewertung von Querschnitten außerörtlicher Straßen [Safety assessment of cross-sections of rural roads]. Berichte der Bundesanstalt für Straßenwesen. Unterreihe Verkehrstechnik. Heft V5*. Bergisch-Gladbach: Bundesanstalt für Straßenwesen (BASt).

Breker, S., Henriksson, P., Eeckhout, G., Falkmer, T., Siren, A., Hakamies-Blomqvist, L., et al. (2002/2003). *Problems of elderly in relation to the driving task and relevant critical scenarios. AGILE Deliverable 1.1.* Retrieved 07.03.2007, from http://www.agile.iao.-fraunhofer.de/downloads/agile_d1_1.pdf

Brewer, F. W. (2002). Mental Models. In L. Nadel (Ed.), *Encyclopedia of cognitive science* (Vol. 3, pp. 1–6). London: Wiley.

Broadbent, D. (1958). *Perception and communication*. London: Pergamon Press.

Brocke, B., Strobel, A., & Müller, J. (2003). Sensation Seeking: Eine biopsychologische Mehr-Ebenen-Theorie. [Sensation seeking: A biopsychological multilevel theory]. In M. Roth & P. Hammelstein (Eds.), *Sensation Seeking – Konzeption, Diagnostik, Anwendung [Sensation seeking - concepts, diagnostics and application]* (pp. 29–51). Göttingen: Hogrefe.

Brocke, B., Tasche, K. G., & Beauducel, A. (1997). Biopsychological foundations of extraversion: Differential effort reactivity and state control. *Personality and Individual Differences, 22*(4), 447–458.

Brosius, F. (2005). *SPSS- Programmierung [SPSS - Programming]*. Bonn: mitp.

Brosius, F. (2008). *SPSS 16*. Heidelberg: mitp.

Broughton, K. L. M., Switzer, F., & Scott, D. (2007). Car following decisions under three visibility conditions and two speeds tested with a driving simulator. *Accident Analysis & Prevention, 39*(1), 106–116.

Brown, I. D. (1978). Dual task methods of assessing workload. *Ergonomics, 21*, 221–224.

Bruce, V., Green, P. R., & Georgeson, M. A. (1996). *Visual perception. Physiology, psychology and ecology* (3rd ed.). Hove: Psychology Press.

Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion [Introduction to the construction of tests and questionnaires]* (2nd ed.). München: Pearson Studium.

Byrne, B. M. (2001). *Structural equation modeling with AMOS. Basic concepts, applications and programming*. Mahwah, NJ: Erlbaum.

Cafiso, S., La Cava, G., & Montella, A. (2007). Safety Index for Evaluation of Two-Lane Rural Highways. *Transportation Research Record: Journal of the Transportation Research Board*(2019), 136-145.

Carsten, O., & Brookhuis, K. (2005). Issues arising from the HASTE experiments. *Transportation Research Part F, 8*(2), 191–196.

Cavallo, V., Brun-Del, M., Laya, O., & Neboit, M. (1988). Perception and anticipation in negotiating curves: The role of driving experience. In A. G. Gale (Ed.), *Vision in Vehicles - II*. Amsterdam: Elsevier.

Cavallo, V., & Cohen, A. S. (2001). Perception. In P.-E. Barjonet (Ed.), *Traffic psychology today* (pp. 63–89). Boston, Dordrecht, London: Kluwer Academic Publisher.

Cavallo, V., & Laurent, M. (1988). Visual information and skill level in time-to-collision estimation. *Perception, 17*(5), 623–632.

Chaparro, A., Wood, J. M., & Carberry, T. (2005). Effects of age and auditory and visual dual tasks on closed-road driving performance. *Optometry and Vision Science, 82*(8), 747–754.

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

Charlton, S. G. (2007). Delineation effects in overtaking lane design. *Transportation Research Part F, 10*, 153–163.

Chipman, M. L., MacGregor, C. G., Smiley, A. M., & Lee-Gosselin, M. (1993). The role of exposure in comparisons of crash risk among different drivers and driving environments. *Accident Analysis & Prevention, 25*(2), 207–211.

Cnossen, F. (2000). *Adaptive strategies and goal management in car drivers. Dissertation: Psychologische, Pedagogische en Sociologische Wetenschappen, Rijksuniversiteit Groningen NL*. Retrieved March 26, 2007, from http://irs.ub.rug.nl/ppn/238175715

Cnossen, F., Meijman, T., & Rothengatter, T. (2004). Adaptive strategy changes as a function of task demands: A study of car drivers. *Ergonomics, 47*(2), 218–236.

Cnossen, F., Rothengatter, T., & Meijman, T. (2000). Strategic changes in task performance in simulated car driving as an adaptive response to task demands. *Transportation Research Part F, 3*, 123–140.

Cohen, A. S. (1987). Blickverhalten des Fahrzeuglenkers als Komponente des Verkehrssicherheitsverhaltens [Car drivers' gaze behaviour as integral part of traffic-safety behaviour]. In H. Häcker (Ed.), *Fortschritte der Verkehrspsychologie [Advances in traffic psychology]* (Vol. 2). Köln: Verlag TÜV Rheinland.

Cohen, A. S. (1997). Menschengerechte Straßenbreite [Human-centered road width]. In B. Schlag (Ed.), *Fortschritte der Verkehrspsychologie 1996 [Advances in traffic psychology 1996]* (pp. 373–382). Bonn: Deutscher Psychologen Verlag.

Collet, C., Petit, C., Priez, A., & Dittmar, A. (2005). Stroop color–word test, arousal, electrodermal activity and performance in a critical driving situation. *Biological Psychology, 69*(2), 195–203.

Comte, S. L., & Jamson, A. H. (2000). Traditional and innovative speed-reducing measures for curves: An investigation of driver behaviour using a driving simulator. *Safety Science, 36*(3), 137–150.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98–104.

Crundall, D., Shenton, C., & Underwood, G. (2004a). Eye movements during active car-following. *Perception, 33*, 575–586.

Crundall, D., Shenton, C., & Underwood, G. (2004b). Eye movements during intentional car following. *Perception, 33*, 975–986.

Crundall, D., Underwood, G., & Chapman, P. (1999). Driving experience and the functional field of view. *Perception, 28*(9), 1075–1087.

Crundall, D., Underwood, G., & Chapman, P. (2002). Attending to the peripheral world while driving. *Applied Cognitive Psychology, 16*(4), 459–475.

Damasio, A. R. (1994). *Descartes' Error: Emotion, reason, and the human brain*. New York: Putnam.

Dawson, M. E., Schell, A. M., & Filion, D. L. (2007). The electrodermal system. In J. T. Cacioppo, L. G. Tassinary & G. G. Berntson (Eds.), *Handbook of psychophysiology* (3rd ed., pp. 159–181). New York: Cambridge University Press.

de Waard, D. (1996). *The Measurement of Drivers' Mental Workload. The Traffic Research Centre VSC, University of Groningen, The Netherlands*. Retrieved March 7, 2007, from http://www.home.zonnet.nl/waard2/mwl.htm

de Waard, D. (2002). Mental workload. In R. Fuller & J. A. Santos (Eds.), *Human Factors for Highway Engineers*. Amsterdam: Pergamon.

de Waard, D., Jessurun, M., Steyvers, F. J. J. M., Raggatt, P. T. F., & Brookhuis, K. A. (1995). Effect of road layout and road environment on driving performance, drivers' physiology and road appreciation. *Ergonomics, 38*(7), 1395–1407.

Derry, S. J. (1996). Cognitive schema theory in the constructivist debate. *Educational Psychologist, 31*(3-4), 163–174.

Dietze, M. (2007). RoadView. Visualisation and analysis software for roads: TU Dresden; Chair of Road Planning and Road Design.

Dietze, M., Ebersbach, D., Lippold, C., Mallschützke, K., & Gatti, G. (2005). *Road Geometry, driving behaviour and road safety. Safety performance functions - Internal Report I. RIPCORD-ISEREST. EU FP 6, Contract No. 50 61 84.* Retrieved July 13, 2008, from http://ripcord.bast.de/pdf/RI-TUD-WP10-R1-Basics.pdf

Dilling, J. (1973). Fahrverhalten von Kraftfahrzeugen auf kurvigen Strecken [Driving behaviour of motor vehicles on curved roads]. In *Straßenbau und Straßenverkehrstechnik, Heft 151*. Bonn: Bundestministerium für Verkehr.

DIN EN ISO 15007-1. (2003). *Messung des Blickverhaltens von Fahrern bei Fahrzeugen mit Fahrerinformations- und -assistenzsystemen. Teil 1: Begriffe und Parameter [Road vehicles - Measurement of driver visual behaviour with respect to transport information and control systems - Part 1: Definitions and parameters]*. Berlin: Beuth Verlag.

Dingus, T. A., Hetrick, S., & Mollenhauer, M. (1999). Empirical methods in support of crash avoidance model building and benefit estimation. *ITS Journal, 5*, 93–125.

Donges, E. (1978). A two-level model of driver steering behavior. *Human Factors, 20*(6), 691–707.

Donges, E. (1999). A conceptual framework for active safety in road traffic. *Vehicle System Dynamics, 32*(2/3), 113–128.

Downing, C. J. (1988). Expectancy and visual-spatial attention: Effects on perceptual quality. *Journal of Experimental Psychology: Human Perception and Performance, 14*(2), 188–202.

Draper, M. H., Viirre, E. S., Furness, T. A., & Gawron, V. J. (2001). Effects of image scale and system time delay on simulator sickness within head-coupled virtual environments. *Human Factors, 43*(1), 129–146.

Driel, C. J. G. v., Davidse, R. J., & Maarseveen, M. F. A. M. v. (2004). The effects of an edgeline on speed and lateral position: A meta-analysis. *Accident Analysis & Prevention, 36*(4), 671–682.

Duh, H. B.-L., Parker, D. E., Philips, J. O., & Furness, T. A. (2004). 'Conflicting' motion cues to the visual and vestibular self-motion systems around 0.06 Hz evoke simulator sickness. *Human Factors, 46*(1), 142–153.

Dulisse, B. (1997). Methodological issues in testing the hypothesis of risk compensation. *Accident Analysis & Prevention, 29*(3), 285–292.

Durth, W. (1974). Ein Beitrag zur Erweiterung des Modells für Fahrer, Fahrzeug und Straße in der Straßenplanung [An article on extending driver, vehicle and road models for road planning]. In *Straßenbau und Straßenverkehrstechnik, Heft 163*. Bonn: Bundesministerium für Verkehr.

Dutta, A., Fisher, D. L., & Noyce, D. A. (2004). Use of a driving simulator to evaluate and optimize factors affecting understandability of variable message signs. *Transportation Research Part F, 7*(4-5), 209–227.

Ebersbach, D. (2006). *Entwurfstechnische Grundlagen für ein Fahrerassistenzsystem zur Unterstützung des Fahrers bei der Wahl seiner Geschwindigkeit [Technical outline of an advanced driver assistance system supporting drivers' speed choice]. Dissertation, TU Dresden*. Retrieved July 25, 2008, from http://nbn-resolving.de/urn:nbn:de:swb:14-1166627418341-24625

Ebersbach, D., & Mayser, C. (2004). Curve-Speed Assistant - Unterstützung des Fahrers bei der Wahl seiner Geschwindigkeit auf Außerortsstraßen. [Curve-Speed Assistant - a driver assistance system to drive the appropriate speed]. In VDI (Ed.), *Integrierte Sicherheit und Fahrerassistenzsysteme [Integrated safety and driver assistance systems]. 21. Internationale VDI/VW-Gemeinschaftstagung. VDI-Berichte 1864* (pp. 497–529). Düsseldorf: VDI Verlag.

Ellinghaus, D., Schlag, B., & Steinbrecher, J. (1990). Leistungsfähigkeit und Fahrverhalten älterer Kraftfahrer [Capability and driving behaviour of the older]. In Bundesanstalt für Straßenwesen (Ed.), *Unfall- und Sicherheitsforschung Straßenverkehr. Heft 80.* Bremerhaven: Wirtschaftsverlag NW.

Ellinghaus, D., & Steinbrecher, J. (2003). *Fahren auf Landstraßen. Traum oder Albtraum? [Driving on rural roads. Dream or nightmare?] Continental AG Hannover (Ed.), UNIROYAL-Verkehrsuntersuchung 28. Hannover: Continental AG. Electronic version*. Retrieved July 5, 2009, from http://www.conti-online.com/generator/www/de-

/de/uniroyal/automobil/themen/05_downloads/02_verkehrsuntersuchungen/download/verkehrsuntersuchung28_de.pdf

Elvik, R. (2002). The importance of confounding in observational before-and-after studies of road safety measures. *Accident Analysis & Prevention, Vol. 34*, 631–635.

Elvik, R. (2006). Laws of accident causation. *Accident Analysis & Prevention, 38*(4), 742–747.

Elvik, R. (2008). Dimensions of road safety problems and their measurement. *Accident Analysis & Prevention, 40*(3), 1200–1210.

Elvik, R., Høye, A., Vaa, T., & Sørensen, M. (2009). *The handbook of road safety measures* (2nd ed.). Bingley, UK: Emerald Group.

Elvik, R., & Vaa, T. (2004). *The handbook of road safety measures*. Amsterdam: Elsevier.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*, 32–64.

Endsley, M. R. (2000). Theoretical underpinnings of situational awareness: A critical review. In M. R. Endsley & D. J. Garland (Eds.), *Situation awareness analysis and measurement* (pp. 3–32). Mahwah, NJ: Lawrence Erlbaum.

Endsley, M. R., & Garland, D. J. (Eds.). (2000). *Situation awareness analysis and measurement*. Mahwah, NJ: Lawrence Erlbaum.

Eriksen, C. W., & St. James, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics, 40*(4), 225–240.

European Commission. (2001). *White Paper: European transport policy for 2010: time to decide*. Luxembourg: Office for Official Publications of the European Communities.

European Commission. (2007). *Commission recommendation of 22 December 2006 on safe and efficient in-vehicle information and communication systems: update of the European Statement of Principles on human machine interface (2007/78/EC)*. Retrieved April 5, 2009, from http://eur-lex.europa.eu/LexUriServ/site/en/oj/2007/l_032/l_03220070206en02000241.pdf

Evans, L. (1986). Risk homeostasis theory and traffic accident data. *Risk Analysis, 6*(1), 81–94.

Evans, L. (2004). *Traffic safety*. Bloomfield Hills, MI: Science Serving Society.

Eysenck, H. J. (1977). *The biological basis of personality* (3rd ed.). Springfield: Thomas.

Eysenck, M. W., & Keane, M. (2005). *Cognitive psychology: A student's handbook* (5th ed.). Hove, East Sussex: Psychology Press.

Fajen, B. R. (2005). Calibration, information, and control strategies for braking to avoid a collision. *Journal of Experimental Psychology: Human Perception and Performance, 31*(3), 480–501.

Farah, H., Yechiam, E., Bekhor, S., Toledo, T., & Polus, A. (2008). Association of risk proneness in overtaking maneuvers with impaired decision making. *Transportation Research Part F, 11*(5), 313–323.

FGSV (Forschungsgesellschaft für Straßen- und Verkehrswesen). (2001). *Merkblatt für die Auswertung von Straßenverkehrsunfällen. Teil 2: Maßnahmen gegen Unfallhäufungen [Leaflet for the analysis of road-traffic accidents. Part 2: Measures against accident accumulations]*. Köln: FGSV-Verlag.

FGSV (Forschungsgesellschaft für Straßen- und Verkehrswesen). (2003). *Empfehlungen für die Sicherheitsanalyse von Straßennetzen (ESN) [Recommendations for analyzing the safety of road networks]*. Köln: FGSV-Verlag.

FGSV (Forschungsgesellschaft für Straßen- und Verkehrswesen) (Ed.). (1995). *Richtlinien für die Anlage von Straßen. Teil: Linienführung (RAS-L) [Guidelines for the construction of roads. Section: Alignment (RAS-L)]*. Köln: FGSV-Verlag.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London: Sage.

Fischer, T., Langner, R., Birbaumer, N., & Brocke, B. (2008). Arousal and attention: Self-chosen stimulation optimizes cortical excitability and minimizes compensatory effort. *Journal of Cognitive Neuroscience, 20*(8), 1–11.

Fitzpatrick, K., Elefteriadou, L., Harwood, D. W., Collins, J. M., McFadden, J., Anderson, I. B., et al. (2000). *Speed prediction for two-lane rural highways. Report FHWA-RD-99-171*. Retrieved March 26, 2007, from http://www.tfhrc.gov/safety/ihsdm/pdfs/99-171.pdf

Fitzpatrick, K., Miaou, S.-P., Brewer, M., Carlson, P., & Wooldridge, M. D. (2005). Exploration of the relationships between operating speed and roadway features on tangent sections. *Journal of Transportation Engineering, 131*(4), 261–269.

Fitzpatrick, K., Wooldridge, M. D., Tsimhoni, O., Collins, J. M., Green, P., Bauer, K. M., et al. (2000). *Alternative design consistency rating methods for two-lane rural highways. Report FHWA-RD-99-172*. Retrieved February 7, 2006, from http://www.-tfhrc.gov/safety/ihsdm/pdfs/99-172.pdf

Flanagan, M. B., May, J. G., & Dobie, T. G. (2004). The role of vection, eye movements and postural instability in the etiology of motion sickness. *Journal of Vestibular Research: Equilibrium & Orientation, 14*(4), 335–346.

Flanagan, M. B., May, J. G., & Dobie, T. G. (2005). Sex differences in tolerance to visually-induced motion sickness. *Aviation, Space, and Environmental Medicine, 76*(7), 642–646.

Franken, R. E. (2007). *Human motivation* (6th ed.). Belmont, CA, USA: Thomson Wadsworth.

Friedel, T. (2005). *Subjektive Kategorisierung von Landstraßen [Subjective categorisation of rural roads].* Unpublished Diploma thesis, TU Dresden; Chair of Traffic and Transportation Psychology, Dresden.

Fuller, R. (1984). A conceptualization of driving behaviour as threat avoidance. *Ergonomics, 27*(11), 1139–1155.

Fuller, R. (2005). Towards a general theory of driver behaviour. *Accident Analysis & Prevention, 37*(3), 461–472.

Fuller, R. (2007). Motivational determinants of control in the driving task. In C. Cacciabue (Ed.), *Modelling driver behaviour in automotive environments* (pp. 165–188). London: Springer.

Fuller, R., McHugh, C., & Pender, S. (2008). Task difficulty and risk in the determination of driver behaviour. *Revue Européenne de Psychologie Appliquée / European Review of Applied Psychology, 58*(1), 13–21.

Fuller, R., & Santos, J. A. (2002). *Human factors for highway engineers*. Amsterdam: Pergamon.

Gabauer, D. J., & Gabler, H. C. (2008). Comparison of roadside crash injury metrics using event data recorders. *Accident Analysis & Prevention, 40*(2), 548–558.

Garson, D. G. (2008). *Logistic Regression*. Retrieved February 13, 2009, from http://faculty.chass.ncsu.edu/garson/PA765/logistic.htm

Gawron, V. J. (2008). *Human performance, workload, and situational awareness measures handbook* (2nd ed.). Boca Raton: CRC Press.

Gelau, C., & Krems, J. F. (2004). The occlusion technique: A procedure to assess the HMI of in-vehicle information and communication systems. *Applied Ergonomics, 35*(3), 185–187.

Gibson, J. J. (1986). *The ecological approach to visual perception*. Hillsdale (NJ): Erlbaum.

Glaser, W. R., Waschulewski, H., & Schmid, D. (2005). I-TSA. Ein standardisiertes Verfahren zur Bewertung der Auswirkungen von Fahrerinformations- und Fahrerassistenzsystemen auf die Verkehrssicherheit [I-TSA - A standardized procedure to assess the safety impact of driver-information and driver-assistance systems]. In VDI-Gesellschaft Fahrzeug- und Verkehrstechnik (Ed.), *Der Fahrer im 21. Jahrhundert. VDI-Berichte, Nr. 1919* (pp. 17–40). Düsseldorf: VDI-Verlag.

Godley, S. T., Triggs, T. J., & Fildes, B. N. (2000). Speed reduction mechanisms of transverse lines. *Transportation Human Factors, 2*(4), 297–312.

Godley, S. T., Triggs, T. J., & Fildes, B. N. (2002). Driving simulator validation for speed research. *Accident Analysis & Prevention, 34*(5), 589–600.

Godley, S. T., Triggs, T. J., & Fildes, B. N. (2004). Perceptual lane width, wide perceptual road centre markings and driving speeds. *Ergonomics, 47*(3), 237–256.

Godthelp, H. (1988). The limits of path error-neglecting in straight lane driving. *Ergonomics, 31*(4), 609–619.

Godthelp, H., Milgram, P., & Blaauw, G. J. (1984). The development of a time-related measure to describe driving strategy. *Human Factors, 26*(3), 257–268.

Goldberg, J. H., & Kotval, X. P. (1999). Computer interface evaluation using eye movements: Methods and constructs. *International Journal of Industrial Ergonomics, 24*(6), 631–645.

Goldenbeld, C., & van Schagen, I. (2007). The credibility of speed limits on 80 km/h rural roads: The effects of road and person(ality) characteristics. *Accident Analysis & Prevention, 39*(6), 1121–1130.

Goldstein, E. B. (2002). *Sensation and perception* (6th ed.). Pacific Grove, CA: Wadsworth.

Goldstein, E. B. (Ed.). (2005). *Blackwell handbook of sensation and perception*. Oxford: Blackwell.

Gordon, C. P. (2009). Crash studies of driver distraction. In M. A. Regan, J. D. Lee & K. L. Young (Eds.), *Driver distraction. Theory, effects, and mitigation* (pp. 281–304). Boca Raton, FL: CRC Press.

Gray, R., & Regan, D. (1999). Do monocular time-to-collision estimates necessarily involve perceived distance? *Perception, 28*(10), 1257–1264.

Gray, R., & Regan, D. (2000). Risky driving behavior: A consequence of motion adaption for visually guided motor action. *Journal of Experimental Psychology: Human Perception and Performance, 26*(6), 1721–1732.

Green, M. (2000). "How long does it take to stop?" Methodological analysis of driver perception-brake times. *Transportation Human Factors, 2*(3), 195–216.

Groeger, J. A. (2000). *Understanding driving. Applying cognitive psychology to a complex everyday task*. Hove, East Sussex: Psychology Press.

Groeger, J. A. (2002). Trafficking in cognition: Applying cognitive psychology to driving. *Transportation Research Part F, 5*(4), 235–248.

Groeger, J. A., & Chapman, P. R. (1996). Judgement of traffic scenes: The role of danger and difficulty. *Applied Cognitive Psychology, 10*(4), 349–364.

Gross, F., & Jovanis, P. P. (2007). Estimation of the safety effectiveness of lane and shoulder width: Case-control approach. *Journal of Transportation Engineering, 133*(6), 362–369.

Gstalter, H., & Fastenmeier, W. (1995). Auswirkungen von Navigationsinformationen im Kraftfahrzeug: Mögliche Sicherheitseffekte, wichtige Fahrervariablen und Folgerungen für empirische Untersuchungen [Effect of navigation information in motor vehicles: Potential safety effects, important driver variables and implications for empirical research]. In W. Fastenmeier (Ed.), *Autofahrer und Verkehrssituation. Neue Wege zur Bewertung von Sicherheit und Zuverlässigkeit moderner Straßenverkehrssysteme [Car driver and traffic situation. New approaches for the evaluation of safety and reliability of modern road-traffic systems]* (pp. 79–95). Köln: Verlag TÜV Rheinland.

Hacker, W. (1994). Moving from cognition to action? Control theory and beyond. *Applied Psychology: An International Review, 43*(3), 379–382.

Hacker, W. (2003). Action Regulation Theory: A practical tool for the design of modern work processes? *European Journal of Work and Organizational Psychology, 12*(1), 105–130.

Hacker, W. (2005). *Allgemeine Arbeitspsychologie. Psychische Regulation von Wissens-, Denk- und körperlicher Arbeit [General work psychology. Psychological regulation of knowledge-, thinking- and physical work]* (2nd ed.). Bern: Hans Huber.

Haight, F. A. (2004). Accident proneness: The history of an idea. In T. Rothengatter & R. D. Huguenin (Eds.), *Traffic and transportation psychology. Proceedings of the ICTTP 2000* (pp. 421–432). Oxford: Elsevier.

Hancock, P. A., Lesch, M., & Simmons, L. (2003). The distraction effects of phone use during a crucial driving maneuver. *Accident Analysis & Prevention, 35*(4), 501–514.

Hancock, P. A., & Verwey, W. B. (1997). Fatigue, workload and adaptive driver systems. *Accident Analysis & Prevention, 29*(4), 495–506.

Hartkopf, G., & Weber, R. (2005). Die neuen Richtlinien für Landstraßen - Ein Schritt zu selbsterklärenden Straßen? [The new guidelines for rural roads - A step towards self-explaining roads?]. *Straßenverkehrstechnik*(1), 10–18.

Harwood, D. W., Council, F. M., Hauer, E., Hughes, W. E., & Vogt, A. (2000). *Prediction of the expected safety performance of rural two-lane highways. Publication No. FHWA-RD-99-207*. Retrieved April 1, 2009, from http://www.tfhrc.gov/safety/pubs/99207.pdf

Haynes, R., Lake, I. R., Kingham, S., Sabel, C. E., Pearce, J., & Barnett, R. (2008). The influence of road curvature on fatal crashes in New Zealand. *Accident Analysis & Prevention, 40*(3), 843–850.

Hebb, D. O. (1955). Drives and the C. N. S. (conceptual nervous system). *Psychological Review, 62*(4), 243–254.

Heger, R., & Weise, G. (1996). Fahrverhalten und Aktivierung von Kraftfahrern als Kriterien zur Bewertung der Linienführung von Außerortsstraßen [Driving behaviour and activation of drivers as criteria for the evaluation of the alignment of rural roads]. In B. Schlag (Ed.), *Fortschritte der Verkehrspsychologie 1996 [Advances in traffic psychology 1996]* (pp. 395–401). Bonn: Deutscher Psychologen Verlag.

Heino, A., van der Molen, H. H., & Wilde, G. J. S. (1996). Differences in risk experience between sensation avoiders and sensation seekers. *Personality and Individual Differences, 20*(1), 71–79.

Herzberg, P. Y., & Schlag, B. (2003). Sensation Seeking und Verhalten im Straßenverkehr [Sensation seeking and behaviour in road traffic]. In M. Roth & P. Hammelstein (Eds.), *Sensation Seeking - Konzeption, Diagnostik und Anwendung [Sensation seeking - concepts, diagnostics and application]* (pp. 162–182). Göttingen: Hogrefe.

Hesketh, B., & Godley, S. T. (2002). A comparison of time estimations in driving with target-only in motion, self-only in motion, and self-and-target in motion. *Ecological Psychology, 14*(3), 111–125.

Hicks, T. G., & Wierwille, W. W. (1979). Comparison of five mental workload assessment procedures in a moving-base driving simulator. *Human Factors, 21*, 129–143.

Hoedemaeker, M., & Brookhuis, K. A. (1998). Behavioural adaptation to driving with an adaptive cruise control (ACC). *Transportation Research Part F, 1*(2), 95–106.

Hoffmann, E. R., & Mortimer, R. G. (1994). Drivers' estimates of time to collision. *Accident Analysis & Prevention, 26*(4), 511–520.

Hoffmann, S., & Buld, S. (2006). Darstellung und Evaluation eines Trainings zum Fahren in der Fahrsimulation. In VDI-Gesellschaft Fahrzeug- und Verkehrstechnik [Driving in a simulator. Design and evaluation of a training programme]. In VDI-Gesellschaft Fahrzeug- und Verkehrstechnik (Ed.), *Integrierte Sicherheit und Fahrerassistenzsysteme (VDI-Berichte, Nr. 1960) [Integrated safety and driver assistance systems]* (pp. 113–132). Düsseldorf: VDI-Verlag.

Horberry, T., Anderson, J., & Regan, M. A. (2006). The possible safety benefits of enhanced road markings: A driving simulator evaluation. *Transportation Research Part F, 9*(1), 77–87.

Horberry, T., Anderson, J., Regan, M. A., Triggs, T. J., & Brown, J. (2006). Driver distraction: The effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance. *Accident Analysis & Prevention, 38*(1), 185–191.

Horrey, W. J., Wickens, C. D., & Consalus, K. P. (2006). Modeling drivers' visual attention allocation while interacting with in-vehicle technologies. *Journal of Experimental Psychology: Applied, 12*(2), 67–78.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York: Wiley.

Howarth, P. A., & Hodder, S. G. (2008). Characteristics of habituation to motion in a virtual environment. *Displays, 29*(2), 117–123.

Hoyos, C., Graf. (1988). Mental load and risk in traffic behaviour. *Ergonomics, 31*(4), 571–584.

Hoyos, C., Graf, & Kastner, M. (1987). Belastung und Beanspruchung von Kraftfahrern [Demand and workload of motor-vehicle drivers]. In Bundesanstalt für Straßenwesen (Ed.), *Unfall- und Sicherheitsforschung Straßenverkehr [Accident- and safety research in road traffic]. Heft 59*. Bremerhaven: Wirtschaftsverlag NW.

Hudson, J. A., & Fivush, R. K., Janet. (1992). Scripts and episodes: The development of event memory. *Applied Cognitive Psychology, 6*(6), 483–505.

IRTAD (International Traffic Safety Data and Analysis Group). (2007). *Fatalities by road location*. Retrieved January 9, 2007, from http://www.cemt.org/irtad/-IRTADPUBLIC/we32.html

ISO 10075-3. (2004). *Ergonomic principles related to mental workload. Part 3: Principles and requirements concerning methods for measuring and assessing mental workload*. Geneve: International Organization for Standardization.

ISO 10075. (1991). *Ergonomic principles related to mental work-load - General terms and definitions*. Geneve: International Organization for Standardization.

Jacob, R. J. K. (1995). Eye tracking in advanced interface design. In W. Barfield & T. A. Furness III (Eds.), *Virtual environments and advanced interface design* (pp. 258–288). Oxford: Oxford University Press.

James, L. R., Mulaik, S. A., & Brett, J. M. (2006). A Tale of two methods. *Organizational Research Methods, 9*(2), 233–244.

Jamson, A. H., & Merat, N. (2005). Surrogate in-vehicle information systems and driver behaviour: Effects of visual and cognitive load in simulated rural driving. *Transportation Research Part F, 8*(2), 79–96.

Janssen, W., & Tenkink, E. (1988). Risk homeostasis theory and its critics: Time for an agreement. *Ergonomics, 31*(4), 429–433.

Johansson, E., Engström, J., Cherri, C., Nodarie, E., Toffetti, A., Schindhelm, R., et al. (2004). *AIDE Deliverable 2.2.1.: Review of existing techniques and metrics for IVIS and ADAS assessment.* Retrieved March 28, 2007, from http://www.aide-eu.org/pdf/-sp2_deliv/aide_d2-2-1.pdf

Johnson-Laird, P. N. (1983, reprint 1990). *Mental models*. Cambridge: Cambridge University Press.

Jonah, B. A. (1997a). Sensation seeking and risky driving. In T. Rothengatter & E. Carbonell Vaya (Eds.), *Traffic and transport psychology - theory and applocation* (pp. 259–268). Amsterdam: Elsevier.

Jonah, B. A. (1997b). Sensation seeking and risky driving: A review and synthesis of the literature. *Accident Analysis & Prevention, 29*(5), 651–665.

Jones, K. S. (2003). What is an affordance? *Ecological Psychology, 15*(2), 107–114.

Jürgensohn, T. (2007). Control theory models of driving behaviour. In C. Cacciabue (Ed.), *Modelling driver behaviour in automotive environments* (pp. 277–292). London: Springer.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.

Kämpfe, B., Schlag, B., & Weller, G. (2005). Streckencharakteristik und Fahrfehler [Road characteristics and driving errors]. *Straßenverkehrstechnik, 2005*(11), 564–571.

Kämpfe, B., Weller, G., & Schlag, B. (2004). Einfluss verschiedener Verkehrswegegestaltungen auf die Verübung von Fahrfehlern [Effect of different road designs on committing driving errors]. In B. Schlag (Ed.), *Verkehrspsychologie. Mobilität - Sicherheit - Fahrerassistenz [Traffic psychology. Mobility - safety - driving assistance]* (pp. 29–46). Lengerich: Pabst.

Kaptein, N., Janssen, W., & Claessens, M. (2002). A study of subjective road categorization and driving behaviour. In R. Fuller & J. A. Santos (Eds.), *Human Factors for Highway Engineers* (pp. 59-76). Amsterdam: Pergamon.

Kaptein, N., Theeuwes, J., & van der Horst, R. (1996). Driving simulator validity: Some considerations. *Transportation Research Record, 1550*, 30–36.

Kastner, S., & Ungerleider, L. G. (2000). Mechanisms of visual attention in the human cortex. *Annual Review of Neuroscience, 23*, 315–341.

Kelly, G. A. (1955/1991). *The psychology of personal constructs*. London: Routledge.

Kennedy, R. S., Drexler, J. M., Compton, D. E., Stanney, K. M., Lanham, D. S., & Harm, D. L. (2003). Configural scoring of simulator sickness, cybersickness, and space adaptation syndrome: similarities and differences. In L. J. Hettinger & M. Haas (Eds.), *Virtual and adaptive environments. Applications, implications, and human performance issues* (pp. 247–278). Mahwah, NJ: Erlbaum.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator Sickness Questionnaire: An enhanced method for quantifying simulator sickness. *International Journal of Aviation Psychology, 3*(3), 203–220.

Kinnear, P. R., & Gray, C. D. (2009). *SPSS 16 made simple*. Hove, East Sussex: Psychology Press.

Klauer, S. G., Dingus, T. A., Neale, V. L., & Sudweeks, J., Ramsey. (2006). *Impact of driver inattention on near-crash/crash risk: An analysis using the 100-Car naturalistic driving study data. NHTSA report DOT HS 810 594*. Retrieved March 20, 2007, from http://www-nrd.nhtsa.dot.gov/departments/nrd-13/driver-distraction/PDF/DriverInattention.pdf

Klebelsberg, D. (1982). *Verkehrspsychologie [Traffic psychology]*. Berlin: Springer.

Klein, R. M. (2000). Inhibition of return. *Trends in Cognitive Sciences, 4*(4), 138–147.

Kluwe, R. H. (2006). Informationsaufnahme und Informationsverarbeitung [Information uptake and information processing]. In B. Zimolong & U. Konradt (Eds.), *Ingenieurpsychologie [Engineering psychology]. Enzyklopädie der Psychologie, Themenbereich D, Serie 3, Band 2 [Encyclopedia of psychology, Subject area D, Series 3, Volume 2]* (pp. 35–70). Göttingen: Hogrefe.

Koyuncu, M., & Amado, S. (2008). Effects of stimulus type, duration and location on priming of road signs: Implications for driving. *Transportation Research Part F, 11*(2), 108–125.

Kramer, A. F., Cassavaugh, N., Horrey, W. J., Becic, E., & Mayhugh, J. L. (2007). Influence of age and proximity warning devices on collision avoidance in simulated driving. *Human Factors, 49*(5), 935–949.

Kramer, A. F., & McCarley, J. S. (2003). Oculomotor behaviour as a reflection of attention and memory processes: Neural mechanisms and applications to human factors. *Theoretical Issues in Ergonomics Science, 4*(1/2), 21–55.

Krems, J. S., Keinath, A., Baumann, M., & Jahn, G. (2004). Die Okklusionsmethode: Ein einfaches und valides Verfahren zur Bewertung der visuellen Beanspruchung von Zweitaufgaben? [The occlusion method: An easy and valid procedure for the assessment of the visual workload of secondary tasks?]. In B. Schlag (Ed.), *Verkehrspsychologie. Mobilität - Sicherheit - Fahrerassistenz [Traffic psychology. Mobility - safety - driving assistance]* (pp. 335–349). Lengerich: Pabst.

LaBerge, D., & Brown, V. (1989). Theory of attentional operations in shape identification. *Psychological Review, 96*(1), 101–124.

Lamm, R., Beck, A., Ruscher, T., Mailaender, T., Cafiso, S., & La Cava, G. (2007). *How to make two-lane rural roads safer. Scientific background and guide for practical application*. Southampton: Witpress.

Lamm, R., Psarianos, B., & Mailaender, T. (1999). *Highway design and traffic safety engineering handbook.* New York: McGraw-Hill.

Land, M. F., & Horwood, J. M. (1995). Which parts of the road guide steering. *Nature, 377*, 339–340.

Land, M. F., & Lee, D. N. (1994). Where we look when we steer. *Nature, 369*, 742–744.

Langan-Fox, J., Code, S., & Langfield-Smith, K. (2000). Team mental models: Techniques, methods, and analytic approaches. *Human Factors, 42*(2), 242–271.

Lazarus, R. S. (1991). *Emotion & adaptation*. New York: Oxford University Press.

Lee, D. N. (1976). A theory of visual control of braking based on information about time-to-collision. *Perception, 5*, 437–459.

Lee, H. C., Cameron, D., & Lee, A. H. (2003). Assessing the driving performance of older adult drivers: On-road versus simulated driving. *Accident Analysis & Prevention, 35*(5), 797–803.

Lenné, M. G., Triggs, T. J., & Redman, J. R. (1997). Time of day variations in driving performance. *Accident Analysis & Prevention, 29*(4), 431–437.

Levi, D. M. (1999). Progress and paradigm shifts in spatial vision over the 20 years of ECVP. *Perception, 28*(12), 1443–1459.

Levy, J., Pashler, H., & Boer, E. (2006). Central interference in driving. *Psychological Science, 17*(3), 228–235.

Lewin, K. (1982). Psychologie der Entwicklung und Erziehung [Psychology of development and education]. In C.-F. Graumann, F. E. Weinert & H. Gundlach (Eds.), *Kurt-Lewin-Werkausgabe. Band 6*. Bern: Hans Huber.

Lewis-Evans, B., & Charlton, S. G. (2006). Explicit and implicit processes in behavioural adaptation to road width. *Accident Analysis & Prevention, 38*(3), 610–617.

Lipphard, D. (2005). "Starenkästen" in Deutschland. Zum Forschungsstand und Verbreitungsgrad von ortsfesten Überwachungsanlagen [Fixed speed monitoring systems in Germany. State of the art and overview of the current locations]. *Zeitschrift für Verkehrssicherheit, 51*(4), 189–193.

Lippold, C. (1997). *Weiterentwicklung ausgewählter Entwurfsgrundlagen von Landstraßen [Further development of selected basic design principles for rural roads]. Dissertation Fachbereich Bauingenieurwesen der Technischen Hochschule Darmstadt*. Darmstadt: Technische Hochschule Darmstadt.

Lippold, C., Dietze, M., Krüger, H.-P., Scheuchenpflug, R., & Mark, C. (2005). *Einfluss der Straßenbepflanzung und Straßenraumgestaltung auf das Verhalten der Verkehrsteilnehmer und auf die Sicherheit im Straßenverkehr an Außerortsstraßen [Effect of roadside vegetation and road-environment design on road-user behaviour and safety of rural roads]. Schlussbericht im Auftrag der Bundesanstalt für Straßenwesen [Final report for the BASt] BASt FE02.217/2002/LGB*. Dresden: TU Dresden; Chair of Road Planning and Road Design.

Lippold, C., & Schulz, R. (2007). Orientierungssichtweite - Definition und Beurteilung [Orientation sight distance -Definition and evaluation]. In Bundesministerium für Verkehr Bau und Stadtentwicklung BMVBS; Abteilung Straßenbau Straßenverkehr (Ed.), *Serie: Forschung Straßenbau und Straßenverkehrstechnik [Series: Research in road construction and road-traffic technique]; Heft 977*. Bremerhaven: Wirtschaftsverlag NW.

Litfin, T., Teichmann, M.-H., & Clement, M. (2000). Beurteilung der Güte explorativer Faktorenanalysen im Marketing [Assessment of the quality of exploratory factor analyses in marketing]. *Wirtschaftswissenschaftliches Studium (WiSt)*(5), 283–286.

Liu, L., Watson, B., & Miyazaki, M. (1999). VR for the elderly: Quantitative and qualitative differences in performance with a driving simulator. *CyberPsychology & Behavior, 2*(6), 567–576.

Locke, E. A. (1994). The emperor is naked. *Applied Psychology: An International Review, 43*(3), 367–370.

Lord, R. G., & Levy, P. E. (1994). Moving from cognition to action: A control theory perspective. *Applied Psychology: An International Review, 43*(3), 335–367.

Luck, S. J., & Vecera, S. P. (2002). Attention. In H. Pashler & S. Yantis (Eds.), *Stevens' handbook of experimental psychology: Vol. 1. Sensation and perception* (3rd ed., pp. 235–286). New York: Wiley.

Luiga, I., & Bachmann, T. (2007). Different effects of the two types of spatial pre-cueing: What precisely is "attention" in Di Lollo's and Enns' substitution masking theory? *Psychological Research, 71*(6), 634–640.

Lynam, D., Castle, J., Scoons, J., Lawson, S. D., Hill, J., & Charman, S. (2007). *EuroRAP II. European Road Assessment Core Programme II. EuroRAP II Technical Report (2005-6)*. Retrieved April 1, 2009, from http://www.eurorap.org/library/pdfs/20070508_-EurorapII_Tech_Report_0506.pdf

Machotka, P. (1980). Daniel Berlyne's contributions to empirical aesthetics. *Motivation and Emotion, 4*(2), 113–121.

Mallschützke, K., Gálvez, I., Gatti, G., van de Leur, M., & Dell' Olio, L. (2006). *RiPCORD – iSEREST. Internal report D9.1. Secondary road safety: Actual knowledge and experiences. (EC Contract No.: 50 61 84, 6th Framework Programme)*. Retrieved July 4, 2008, from http://ripcord.bast.de/pdf/RI-INEC-WP9-R1-State_of_the_Art.pdf

Manser, M. P., & Hancock, P. A. (2007). The influence of perceptual speed regulation on speed perception, choice, and control: Tunnel wall characteristics and influences. *Accident Analysis & Prevention, 39*(1), 69–78.

Manzey, D. (1998). Psychophysiologie mentaler Beanspruchung [Psychophysiology of mental workload]. In F. Rösler (Ed.), *Ergebnisse und Anwendungen der Psychophysiologie [Results and applications of psychophysiology]. Enzyklopädie der Psychologie, Themenbereich C, Serie I, Band 5 [Encyclopedia of psychology, Subject area C, Series I, Volume 5]* (pp. 799–864). Göttingen: Hogrefe.

Martens, M. H., Comte, S., & Kaptein, N. (1997). *The effects of road design on speed behaviour - A literature review. MASTER Deliverable D1 (Report 2.3.1). VTT*. Retrieved June 27, 2005, from http://www.vtt.fi/rte/projects/yki6/master/rep231.pdf

Martens, M. H., & Fox, M. R. J. (2007). Do familiarity and expectations change perception? Drivers' glances and response to changes. *Transportation Research Part F, 10*(6), 476–492.

Matena, S., Louwerse, R., Schermers, G., Vaneerdewegh, P., Pokorny, P., Gaitanidou, L., et al. (2007). *RiPCORD – iSEREST. Deliverable D3. Road design and environment. Best practice on self-explaining and forgiving roads. (EC Contract No.: 50 61 84, 6th Framework Programme)*. Retrieved July 4, 2008, from http://ripcord.bast.de/pdf/-RIPCORD-ISEREST-Deliverable-D3-Final.pdf

Matena, S., Weber, R., Louwerse, R., Drolenga, H., Vaneerdewegh, P., Pokorny, P., et al. (2006). *RiPCORD – iSEREST. Internal report D3.1. Report on road classification, design of self explaining roads, roadside environment and traffic regulations. (EC Contract No.: 50 61 84, 6th Framework Programme).* Retrieved July 4, 2008, from http://ripcord.bast.de/pdf/RI-BASt-WP3-R1-Road_Categorisation_and_SER.pdf

Maycock, G. (1997). Accident liability: The human perspective. In T. Rothengatter & E. Carbonell Vaya (Eds.), *Traffic & transport psychology: Theory and application* (pp. 65–76). Amsterdam: Pergamon.

McEvoy, S. P., Stevenson, M. R., & Woodward, M. (2007). The prevalence of, and factors associated with, serious crashes involving a distracting activity. *Accident Analysis & Prevention, 39*, 475–482.

McGehee, D. V., Lee, J. D., Rizzo, M., Dawson, J., & Bateman, K. (2004). Quantitative analysis of steering adaptation on a high performance fixed-base driving simulator. *Transportation Research Part F, 7*(3), 181–196.

McKenna, F. P. (1988). What role should the concept of risk play in theories of accident involvement? *Ergonomics, 31*(4), 469–484.

Messer, C. J. (1980). Methodology for evaluating geometric design consistency. *Transportation Research Record*(757), 7–14.

Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance: I. Basic mechanisms. *Psychological Review, 104*(1), 3–65.

Michon, J. A. (1985). A critical view of driver behavior models: What do we know, what should we do? In L. Evans & R. C. Schwing (Eds.), *Human behavior and traffic safety* (pp. 485–524). New York: Plenum Press.

Milleville-Pennel, I., Hoc, J.-M., & Jolly, E. (2007). The use of hazard road signs to improve the perception of severe bends. *Accident Analysis & Prevention, 39*(4), 721–730.

Miura, T. (1990). Active function of eye movement and useful field of view in a realistic setting. In R. Groner, d'Ydewalle & R. Parham (Eds.), *From eye to mind: Information acquisition in perception, search, and reading* (pp. 119–127). Amsterdam: North-Holland.

Mourant, R. R., Ahmad, N., Jaeger, B. K., & Lin, Y. (2007). Optic flow and geometric field of view in a driving simulator display. *Displays, 28*(3), 145–149.

Müller, H. J., & Krummenacher, J. (2002). Aufmerksamkeit [Attention]. In J. Müsseler & W. Prinz (Eds.), *Allgemeine Psychologie [General psychology]* (pp. 118–177). Heidelberg: Spektrum.

Müsseler, J. (2002). Visuelle Wahrnehmung [Visual perception]. In J. Müsseler & W. Prinz (Eds.), *Allgemeine Psychologie [General psychology]* (pp. 14–65). Heidelberg: Spektrum.

Myers, R. S., Ball, K., & Kalina, T. D. (2000). Relation of useful field of view and other screening tests to on-road driving performance. *Perceptual and Motor Skills, 91*(1), 279–290.

Näätänen, R., & Summala, H. (1974). A model for the role of motivational factors in drivers' decision-making. *Accident Analysis & Prevention, 6*(3-4), 243–261.

Näätänen, R., & Summala, H. (1976). *Road-user behavior and traffic accidents*. Amsterdam: North-Holland.

Navon, D., & Miller, J. (2002). Queuing or sharing? A critical evaluation of the single-bottleneck notion. *Cognitive Psychology, 44*(3), 193–251.

Neiss, R. (1988). Reconceptualizing arousal: Psychobiological states in motor performance. *Psychological Bulletin, 103*(3), 345–366.

Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology*. San Francisco: Freeman.

Neuschatz, J. S., Lampinen, J., & Preston, E. L. (2002). The effect of memory schemata on memory and the phenomenological experience of naturalistic situations. *Applied Cognitive Psychology, 16*(6), 687–708.

Nirschl, G., Böttcher, S., Schlag, B., & Weller, G. (2004). Verfahren zur Bewertung der Verkehrssicherheit von Fahrerassistenzsystemen durch objektive Erfassung von Fahrfehlerrisiken [Procedure for evaluating traffic safety of driver assistance systems by objective measurement of driving error risks]. In VDI (Ed.), *Integrierte Sicherheit und Fahrerassistenzsysteme [Integrated safety and driver assistance systems]. 21. Internationale VDI/VW-Gemeinschaftstagung. VDI-Berichte 1864* (pp. 397–420). Düsseldorf: VDI Verlag.

Noland, R. B. (2003). Traffic fatalities and injuries: The effect of changes in infrastructure and other trends. *Accident Analysis & Prevention, 35*(4), 599–611.

Norman, D. A. (1981). Categorization of action slips. *Psychological Review, 88*(1), 1–15.

Norman, D. A. (1998). *The Design of everyday things*. London: MIT Press.

OECD. (1990). *Behavioural adaptations to changes in the road transport system*. Paris: OECD.

OECD. (1999). *Safety strategies for rural roads*. Paris: OECD.

OECD. (2006). *Young Drivers. The Road to Safety*. Paris: OECD.

Olson, J. R., & Biolsi, K. J. (1991). Techniques for representing expert knowledge. In K. A. Ericcson & J. Smith (Eds.), *Toward a general theory of expertise* (pp. 240–285). Cambridge: Cambridge University Press.

Owens, D. A., Wood, J. M., & Owens, J. M. (2007). Effects of age and illumination on night driving: A road test. *Human Factors, 49*(6), 1115–1131.

Park, J.-R., Lim, D.-W., Lee, S.-Y., Lee, H.-W., Choi, M.-H., & Chung, S.-C. (2008). Long-term study of simulator sickness: Differences in EEG response due to individual sensitivity. *International Journal of Neuroscience, 118*(6), 857–865.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin, 116*(2), 220–244.

Pashler, H. (1998). *The psychology of attention*. Cambridge, MA: MIT Press.

Petermann, I. (2006). *Visuelle Wahrnehmung und subjektive Einschätzung von Kurven und von deren Annäherungsbereichen. Der Beitrag psychologischer Einflussgrößen zur Erklärung des Unfallgeschehens auf Außerortsstraßen [Visual perception and subjective evaluation of curves and their approach zones. The contribution of psychological variables in explaining accident occurrence on rural roads].* Unpublished Diploma thesis, TU Dresden; Chair of Traffic and Transportation Psychology, Dresden.

Petermann, I., Weller, G., & Schlag, B. (2007). Beitrag des visuellen Eindrucks zur Erklärung des Unfallgeschehens in Landstraßen [The contribution of the visual Impression to explaining curve accidents on rural roads]. *Straßenverkehrstechnik, 51*(8), 401–408.

Peterson, M. S., Kramer, A. F., Wang, R. F., Irwin, D. E., & McCarley, J. S. (2001). Visual search has memory. *Psychological Science, 12*(4), 287–292.

Pett, M. A., Lackey, N. R., & Sullivan, J. J. (2003). *Making sense of factor analysis : The use of factor analysis for instrument development in health care research*. Thousand Oaks, CA: Sage.

Pew, R. W. (1979). Secondary tasks and workload measurement. In N. Moray (Ed.), *Mental workload. Its theory and measurement* (pp. 23–28). New York: Plenum Press.

Piechulla, W. (2006). Messung und Kontrolle mentaler Fahrerbeanspruchung [Measurement and control of drivers' mental workload]. In U. W. Ebner-Priemer (Ed.), *Ambulantes psychophysiologisches Monitoring - neue Perspektiven und Anwendungen [Ambulant psychophysiological monitoring - new perspectives and applications]* (pp. 21–51). Frankfurt am Main: Peter Lang.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32*(1), 3–25.

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General, 109*(2), 160–174.

Pribram, K. H., & McGuinness, D. (1975). Arousal, activation, and effort in the control of attention. *Psychological Review, 82*(2), 116–149.

PTV Planung Transport Verkehr. (2006). EUSKA Accident Analysis Software: http://www.ptv.de.

Rammin, C. (2006). *Die Anwendung der Road Environment Construct List (RECL) zur subjektiven Bewertung von Landstraßen [Application of the Road Environment Construct List (RECL) for the subjective assessment of rural roads].* Unpublished Student Paper, TU Dresden; Chair of Traffic and Transportation Psychology, Dresden.

Ranney, T. A. (1994). Models of driving behavior: A review of their evolution. *Accident Analysis & Prevention, 26*(6), 733–750.

Rantanen, E. M., & Goldberg, J. H. (1999). The effect of mental workload on the visual field size and shape. *Ergonomics, 42*(6), 816–834.

Räsänen, M., Lajunen, T., Alticafarbay, F., & Aydin, C. (2007). Pedestrian self-reports of factors influencing the use of pedestrian bridges. *Accident Analysis & Prevention, 39*(5), 969–973.

Räsänen, M., & Summala, H. (2000). Car drivers' adjustments to cyclists at roundabouts. *Transportation Human Factors, 2*(1), 1–17.

Rasmussen, J. (1986). *Information processing and human-machine interaction. An approach to cognitive engineering*. New York: North-Holland.

Reason, J. T. (1978). Motion sickness adaptation: A neural mismatch model. *Journal of the Royal Society of Medicine, 71*(11), 819–829.

Reason, J. T. (1990). *Human Error*. Cambridge: Camebridge University Press.

Recarte, M. A., & Nunes, L. M. (2000). Effects of verbal and spatial-imagery tasks on eye fixations while driving. *Journal of Experimental Psychology: Applied, 6*(1), 31–43.

Reed, M. P., & Green, P. A. (1999). Comparison of driving performance on-road and in a low-cost simulator using a concurrent telephone dialling task. *Ergonomics, 42*, 1015–1037.

Reichart, G. (2001). *Menschliche Zuverlässigkeit beim Führen von Kraftfahrzeugen [Human reliability when driving a vehicle]. Forschritt-Berichte VDI, Reihe 22, Nr. 7*. Düsseldorf: VDI Verlag.

Reimer, B., D'Ambrosio, L. A., Coughlin, J. F., Kafrissen, M. E., & Biederman, J. (2006). Using self-reported data to assess the validity of driving simulation data. *Behavior Research Methods & Instrumentation, 38*(2), 314–324.

Reurings, M., Janssen, T., Eenigk, R., Elvik, R., Cardoso, J., & Stefan, C. (2005). *RiPCORD – iSEREST. Internal report D2.1. Accident prediction models and road safety impact assessment: A state-of-the-art. Internal report (EC: Contract no.: 50 61 84, 6th Framework Programme).* Retrieved July 5, 2008, from http://ripcord.bast.de/pdf/RI-SWOV-WP2-R1-State_of_the_Art.pdf

Reymond, G., Kemeny, A., & Droulez, J. (2001). Role of lateral acceleration in curve driving: Driver model and experiments on a real vehicle and a driving simulator. *Human Factors, 43*(3), 483–495.

Richter, P., Debitz, U., & Schulze, F. (2002). Diagnostik von Arbeitsanforderungen und kumulativen Beanspruchungsfolgen am Beispiel eines Call Centers [Diagnostic of job demands and cumulating strain consequences in call center jobs]. *Zeitschrift für Arbeitswissenschaft*(1-2), 67–76.

Richter, P., & Hacker, W. (1998). *Belastung und Beanspruchung. Streß, Ermüdung und Burnout im Arbeitsleben [Demand and workload. Stress, fatigue and burn-out in working life]*. Heidelberg: Asanger.

Richter, P., Wagner, T., Heger, R., & Weise, G. (1998). Psychophysiological analysis of mental load during driving on rural roads – a quasi-experimental field study. *Ergonomics, 41*(5), 593–609.

Richter, P., Weise, G., Wagner, T., & Heger, R. (1996). *Fahrverhalten und psychophysiologische Aktivierung von Kraftfahrern als Bewertungskriterien der Gestaltungsgüte von Straßenverkehrsanlagen. Schlussbericht zum Forschungsprojekt Ri 671/2-1 im Auftrag der Deutschen Forschungsgemeinschaft DFG [Driving behaviour and psychophysiological activation as evaluation criterion for the design quality of roads. Final report]*. Dresden: Technische Universität Dresden.

Riemersma, J. B. J. (1988). An empirical study of subjective road categorization. *Ergonomics, 31*(4), 621–630.

Robbins, T. W. (1997). Arousal systems and attentional processes. *Biological Psychology, 45*(1-3), 57–71.

Rock, P. B., Harris, M. G., & Yates, T. (2006). A Test of the tau-dot hypothesis of braking control in the real world. *Journal of Experimental Psychology: Human Perception and Performance, 32*(6), 1479–1484.

Rogé, J., Otmani, S., Pébayle, T., & Muzet, A. (2008). The impact of age on useful visual field deterioration and risk evaluation in a simulated driving task. *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology, 58*(1), 5–12.

Rohrbaugh, J. W. (1984). The orienting reflex: Performance and central nervous system manifestations. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention*. London: Academic Press.

Rosey, F., Auberlet, J.-M., Bertrand, J., & Plainchault, P. (2008). Impact of perceptual treatments on lateral control during driving on crest vertical curves: A driving simulator study. *Accident Analysis & Prevention, 40*(4), 1513–1523.

Rothengatter, T. (1988). Risk and the absence of pleasure: A motivational approach to modelling road user behaviour. *Ergonomics, 31*(4), 599–607.

Rothengatter, T. (2002). Drivers' illusions – No more risk. *Transportation Research Part F, 5*(4), 249–258.

Rötting, M. (2001). *Parametersystematik der Augen- und Blickbewegungen für arbeitswissenschaftliche Untersuchungen [Taxonomy of eye- and gaze-movement parameters for empirical research in ergonomics]*. Aachen: Shaker.

Rudolf, M., & Müller, J. (2004). *Multivariate Verfahren. Eine praxisorientierte Einführung mit Anwendungsbeispielen in SPSS [Multivariate analyses. A practice-oriented introduction using SPSS]*. Göttingen: Hogrefe.

Rumar, K. (1985). The role of perceptual and cognitive filters in observed behaviour. In L. Evans & R. C. Sching (Eds.), *Human Behaviour and Traffic Safety* (pp. 151–170). New York: Plenum Press.

SafetyNet. (2007). *Annual statistical report 2006 (SafetyNet: EC-Contract No.: TREN-04-FP6TRSI2.395465/506723, Project Number: 1.3.2),* from http://www.erso.eu/-safetynet/fixed/WP1/2006/SN_1_3_ASR_2006_final.pdf

Salvucci, D. D. (2006). Modeling driver behavior in a cognitive architecture. *Human Factors, 48*(2), 362–380.

Salvucci, D. D., & Goldberg, J. H. (2000). *Identifying fixations and saccades in eye-tracking protocols. In Proceedings of the eye tracking research & application symposium 2000. New York: ACM Press (pp. 71–78) electronic version*. Retrieved 24 July 2009, from http://www.cs.drexel.edu/~salvucci/publications/Salvucci-ETRA00.pdf

Sanders, A. F. (1983). Towards a model of stress and human performance. *Acta Psychologica, 53*(1), 61–97.

Sanders, A. F. (1997). A summary of resource theories from a behavioral perspective. *Biological Psychology, 45*(1-3), 5–18.

Sarris, V., & Reiß, S. (2005). *Kurzer Leitfaden der Experimentalpsychologie [Short manual of experimental psychology]*. München: Pearson Studium.

Schade, F.-D. (2000). Verkehrsauffälligkeit von PKW-Fahrern und ihre Entwicklung mit dem Lebensalter - ein Modell [Conspicuous behavior of car drivers and their changes with age - a model]. *Zeitschrift für Verkehrssicherheit, 46*(1), 9–18.

Schandry, R. (1998). *Lehrbuch Psychophysiologie. Körperliche Indikatoren psychischen Geschehens [Textbook psychophysiology. Physiological indicators of psychological processes]* (4th ed.). Weinheim: Beltz.

Schank, R. C., & Abelson, R. (1977). *Scripts, plans, goals and understanding*. Hillsdale, NJ: Erlbaum.

Schlag, B. (1993). Elderly drivers in Germany — Fitness and driving behavior. *Accident Analysis & Prevention, 25*(1), 47–55.

Schlag, B. (1994). Risikoverhalten im Straßenverkehr [Risk behaviour in road traffic]. In A. Flade (Ed.), *Mobilitätsverhalten* (pp. 131–138). Weinheim: Beltz / PVU.

Schlag, B. (2004). *Lern- und Leistungsmotivation [Learning and achievement motivation]* (2nd ed.). Wiesbaden: VS Verlag.

Schlag, B. (2008a). Älter werden und Auto fahren. *Report Psychologie, 33*(2), 75-85.

Schlag, B. (2008b). Einleitung: Wie sicher sind die Älteren im Straßenverkehr? In B. Schlag (Ed.), *Leistungsfähigkeit und Mobilität im Alter* (pp. 19-36). Köln: TÜV Media & Eugen-Otto-Butz Stiftung.

Schlag, B. (Ed.). (2008c). *Leistungsfähigkeit und Mobilität im Alter* (Vol. 3, 'Mobilität und Alter'; Eugen-Otto-Butz Stiftung). Köln: TÜV Media.

Schlag, B., Petermann, I., Weller, G., & Schulze, C. (2009). *Mehr Licht - Mehr Sicht - Mehr Sicherheit? Zum Einfluss unterschiedlicher Licht- und Sichtverhältnisse auf das Fahrerverhalten [More light - more sight - more safety? The influence of different light and sight conditions on driving behaviour]*. Wiesbaden: Verlag für Sozialwissenschaften, VS Research.

Schlender, D. (2008). Simulatorkrankheit in Fahrsimulatoren [Simulator sickness in driving simulators]. *Zeitschrift für Verkehrssicherheit, 54*(2), 74-80.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*(1), 1–66.

Schulz, R. (2007). WatchOut. Eye-movement analysis tool. Dresden: TU Dresden; Chair of Road Planning and Road Design.

Schumacker, R. E., & Lomax, R. G. (2004). *Beginner's guide to structural equation modeling* (2nd ed.). Mahwah, NJ: Erlbaum.

Schwarz, S., & Schlichter, H. G. (2004). Mehr Sicherheit und Komfort in Fahrerassistenzsystemen, Teil 1 und Teil 2 (Sonderdruck, Heft 2, S. 14–20 und Heft 3, S10–21) [More safety and comfort of driver assistance systems, Part 1 and 2 (offprint, issue 2, pp. 14–20 and issue 3, pp. 10–21)]. *Straßen- und Tiefbau, 58*.

Schwebel, D. C., Gaines, J., & Severson, J. (2008). Validation of virtual reality as a tool to understand and prevent child pedestrian injury. *Accident Analysis & Prevention, 40*(4), 1394–1400.

Schweigert, M. (2003). *Fahrerblickverhalten und Nebenaufgaben [Drivers' gaze behaviour and secondary tasks]. Doctoral dissertation. Technische Universität München, Chair of Ergonomics*. Retrieved July 3, 2007, from http://tumb1.biblio.tu-muenchen.de/publ/diss/mw/2003/schweigert.pdf

Shevlina, M., Miles, J. N. V., Davies, M. N. O., & Walke, S. (2000). Coefficient alpha: a useful indicator of reliability? *Personality and Individual Differences, 28*(2), 229–237.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review, 84*(2), 127–190.

Shinar, D. (1977). Curve perception and accidents on curves: An illusive curve phenomenon? *Zeitschrift für Verkehrssicherheit, 23*(1), 16–21.

Shinar, D., Rockwell, T. H., & Malecki, J. A. (1980). The effects of changes in driver perception on rural curve negotiation. *Ergonomics, 23*(3), 263–275.

Shinar, D., Tractinsky, N., & Compton, R. (2005). Effects of practice, age, and task demands, on interference from a phone task while driving. *Accident Analysis & Prevention, 37*(2), 315–326.

Sidaway, B., Fairweather, M., & Sekiya, H. (1996). Time-to-collision estimation in a simulated driving task. *Human Factors, 38*(1), 101–113.

Simons, D. J. (2000). Attentional capture and inattentional blindness. *Trends in Cognitive Sciences, 4*(4), 147–155.

Sivak, M. (1996). The information that drivers use: Is it indeed 90% visual? *Perception, 25*(9), 1081–1089.

Smiley, A. (1999). *Driver speed estimation: What road designers should know. Transportation Research Board , 78th Annual Meeting Workshop on Role of Geometric Design & Human Factors in Setting Speed*. Retrieved March 26, 2007, from http://www.hfn.ca/driver.htm

Sossoumihen, A. J. (2001). *Entwicklung eines Rahmenkonzeptes zur Bewertung der Linienführung von Außerortsstraßen nach der Zielfunktion Fahrsicherheit [Development of a conceptual framework for the assessment of rural-road alignment regarding driving safety].*

Unpublished doctoral dissertation, TU Dresden, Chair of Road Planning and Road Design, Dresden.

Spacek, P. (2005). Track behavior in curve areas: Attempt at typology. *Journal of Transportation Engineering, 131*(9), 669–676.

SPSS 16 Tutorial. *SPSSInc\SPSS16DE\help\main\spssmain.chm*: SPSS Inc.

Statistisches Bundesamt. (2006). *Verkehr. Verkehrsunfälle. 2005. [Traffic. Accidents. 2005] (Fachserie 8, Reihe 7)*. Wiesbaden: Statistisches Bundesamt.

Statistisches Bundesamt. (2007). *Verkehr. Verkehrsunfälle. 2006. [Traffic. Accidents. 2006] (Fachserie 8, Reihe 7)*. Wiesbaden: Statistisches Bundesamt.

Statistisches Bundesamt. (2009). *Verkehr. Verkehrsunfälle. 2008. [Traffic. Accidents. 2008] (Fachserie 8, Reihe 7)*. Wiesbaden: Statistisches Bundesamt.

Steyer, R. (2004). *Beitrag zur Erhöhung der Fahrsicherheit in Kurvenbereichen zweistreifiger Außerortsstraßen [Contribution to increasing driving safety in curved sections of two-lane rural roads].* Unpublished doctoral dissertation, TU Dresden, Chair of Road Planning and Road Design, Dresden.

Steyvers, F. J. J. M. (1993). The measurement of road environment appreciation with a multi-scale construct list. In A.G. Gale, I. D. Brown, C. M. Haslegrave, H. W. Kruysse & S. P. Taylor (Eds.), *Vision in Vehicles-IV* (pp. 203–212). Amsterdam: North-Holland.

Steyvers, F. J. J. M. (1998). Categorisation and appraisal of rural two-lane undivided 80-km/h roads. In A.G.Gale, I.D. Brown, C.M. Haslegrave & S. P. Taylor (Eds.), *Vision in vehicles - VI* (pp. 271–278). Amsterdam: Elsevier.

Steyvers, F. J. J. M., & de Waard, D. (2000). Road-edge delineation in rural areas: Effects on driving behaviour. *Ergonomics, 43*(2), 223–238.

Steyvers, F. J. J. M., Dekker, K., Brookhuis, K. A., & Jackson, A. E. (1994). The experience of road environments under two lighting and traffic conditions: Application of a Road Environment Construct List. *Applied Cognitive Psychology, 8*(5), 497–511.

Stutts, J., Feaganes, J., Reinfurt, D., Rodgmana, E., Hamlett, C., Gish, K., et al. (2005). Driver's exposure to distractions in their natural driving environment. *Accident Analysis & Prevention, 37*(6), 1093–1101.

Summala, H. (1988). Risk control is not adjustment: The zero-risk theory of driver behaviour and its implications. *Ergonomics, 31*(4), 491–506.

Summala, H. (1996). Accident risk and driver behaviour. *Safety Science, 22*(1-3), 103–117.

Summala, H. (1997). Hierarchical model of behavioural adaptation and traffic accidents. In T. Rothengatter & E. Carbonell Vaya (Eds.), *Traffic & transport psychology: Theory and application*. Amsterdam: Pergamon.

Summala, H. (1998). Forced peripheral vision driving paradigma: Evidence for the hypothesis that car drivers learn to keep in lane with peripheral vision. In A. G. Gale, I. D. Brown, C. M. Haslegrave & S. P. Taylor (Eds.), *Vision in Vehicles - VI* (pp. 51–60). Amsterdam: Elsevier.

Summala, H. (2005). Traffic psychology theories: Towards understanding driving behaviour and safety efforts. In G. Underwood (Ed.), *Traffic & transport psychology: Theory and application* (pp. 383–398). Amsterdam: Elsevier.

Summala, H. (2007). Towards understanding motivational and emotional factors in driver behaviour: Comfort through satisficing. In C. Cacciabue (Ed.), *Modelling driver behaviour in automotive environments* (pp. 189–207). London: Springer.

Summala, H., Nieminen, T., & Punto, M. (1996). Maintaining lane position with peripheral vision during in-vehicle tasks. *Human Factors, 38*(3), 442–451.

Taubman-Ben-Ari, O. (2008). Motivational sources of driving and their associations with reckless driving cognitions and behavior. *Revue europeenne de psychologie appliquee, 58*(1), 51–64.

Taylor, D. H. (1964). Drivers' galvanic skin response and the risk of accident. *Ergonomics, 7*, 439–451.

Taylor, M. M. (1999). Perceptual control theory and its application (Editorial). *International Journal of Human-Computer Studies, 50*(6), 433–444.

Theeuwes, J. (1998). Self-explaining roads: Subjective categorisation of road environments. In A.G.Gale, I.D. Brown, C. M. Haslegrave & S. P. Taylor (Eds.), *Vision in vehicles - VI* (pp. 279–287). Amsterdam: Elsevier.

Theeuwes, J. (2000). Commentary on Räsänen and Summala, 'Car drivers' adjustments to cyclists at roundabouts'. *Transportation Human Factors, 2*(1), 19–22.

Theeuwes, J. (2004). Top-down search strategies cannot override attentional capture. *Psychonomic Bulletin & Review, 11*(1), 65–70.

Theeuwes, J., & Godthelp, H. (1995). Self-explaining roads. *Safety Science, 19*, 217–225.

Törnros, J. (1998). Driving behaviour in a real and a simulated road tunnel - a validation study. *Accident Analysis & Prevention, 30*(4), 497–503.

Treat, J. R., Tumbas, N. S., McDonald, S. T., Shinar, D., Hume, R. D., Mayer, R. E., et al. (1977). *Tri-level study of the causes of traffic accidents. Volume I: Casual factor tabulations and assessment. Final report* (No. DOT-HS-034-3-534). Washington: National Highway Traffic Safety Administration.

Tresilian, J. R. (1991). Empirical and theoretical issues in the perception of time to contact. *Journal of Experimental Psychology: Human Perception and Performance, 17*(3), 865–876.

Tresilian, J. R. (1999). Visually timed action: time-out for 'tau'? *Trends in Cognitive Sciences, 3*(8), 301–310.

Trick, L. M., Enns, J. T., Mills, J., & Vavrik, J. (2004). Paying attention behind the wheel: A framework for studying the role of attention in driving. *Theoretical Issues in Ergonomics Science, 5*(5), 385–424.

Tsang, P. S., & Vidulich, M. A. (2006). Mental workload and situation awareness. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (3rd ed., pp. 243–268). Hoboken, NJ: Wiley.

Tsimhoni, O., & Green, P. A. (1999). *Visual Demand of Driving Curves as Determined by Visual Occlusion. Paper presented at the Vision in Vehicles VIII Conference, Boston, MA.* Retrieved December 15, 2006, from http://www.umich.edu/~driving/-publications/VIV-Tsimhoni1999.pdf

Tynan, P. D., & Sekuler, R. (1982). Motion processing in peripheral vision: Reaction time and perceived velocity. *Vision Research, 22*(1), 61–68.

Underwood, G., Chapman, P., & Brocklehurst, N. (2003). Visual attention while driving: Sequences of eye fixations made by experienced and novice drivers. *Ergonomics, 46*(6), 629–646.

Vaa, T. (2007). Modelling driver behaviour on basis of emotions and feelings: Intelligent transport systems and behavioural adaptations. In C. Cacciabue (Ed.), *Modelling driver behaviour in automotive environments* (pp. 208–232). London: Springer.

VaezMousavi, S. M., Barry, R. J., & Rushby, J. A. (2007). Evidence for differentiation of arousal and activation in normal adults. *Acta Neurobiologiae Experimentalis, 67*(2), 179–186.

van Winsum, W., & Godthelp, H. (1996). Speed choice and steering behavior in curve driving. *Human Factors, 38*(3), 434–441.

Velichkovsky, B. M., Rothert, A., Kopf, M., Dornhöfer, S. M., & Joos, M. (2002). Towards an express-diagnostics for level of processing and hazard perception. *Transportation Research Part F, 5*(2), 145–156.

Verwey, W. B. (2000). On-line driver workload estimation. Effects of road situation and age on secondary task measures. *Ergonomics, 43*(2), 187–209.

Verwey, W. B., & Veltman, H. A. (1996). Detecting short periods of elevated workload: A comparison of nine workload assessment techniques. *Journal of Experimental Psychology: Applied, 2*(3), 270–285.

Verwey, W. B., & Zaidel, D. M. (2000). Predicting drowsiness accidents from personal attributes, eye blinks and ongoing driving behaviour. *Personality and Individual Differences, 28*(1), 123–142.

Victor, T. W., Harbluk, J. L., & Engström, J. A. (2005). Sensitivity of eye-movement measures to in-vehicle task difficulty. *Transportation Research Part F, 8*(2), 167–190.

Vogt, A., & Bared, J. G. (1998). *Accident models for two-lane rural roads: Segments and intersections. Report No. FHWA-RD-98-133*. Retrieved April 1, 2009, from http://www.tfhrc.gov/safety/98133/index.html

Voigt, J. (2007). *Der Einfluss der Straßengestaltung auf das Fahrverhalten unter Berücksichtigung der subjektiven Einschätzung der einzelnen Gestaltungselemente [The influence of road design on driving behaviour considering the subjective evaluation of the single design elements].* Unpublished diploma thesis, TU Dresden, Chair of Traffic and Transportation Psychology, Dresden.

Vollrath, M., Meilinger, T., & Krüger, H.-P. (2002). How the presence of passengers influences the risk of a collision with another vehicle. *Accident Analysis & Prevention, 34*(5), 649–654.

von Campenhausen, C. (1993). *Die Sinne des Menschen. Einführung in die Psychophysik der Wahrnehmung [The human senses. Introduction to the psychophysics of perception]* (2nd ed.). Stuttgart: Georg Thieme Verlag.

Wagner, T. (2000). *Die Integration von anforderungsanalytischen und biopsychologischen Methoden zur Bewertung von Arbeitssystemen - eine Felduntersuchung zu Fahraufgaben auf Außerortsstraßen [The integration of demand-analytic and biopsychological methods into the evaluation of work systems - a field study of the driving task on rural roads].*

Unpublished doctoral dissertation, Technische Universität Dresden, Chair of Work and Organizational Psychology, Dresden.

Walker, I., & Brosnan, M. (2007). Drivers' gaze fixations during judgements about a bicyclist's intentions. *Transportation Research Part F, 10*(2), 90–98.

Wallis, G., Chatziastros, A., Tresilian, J., & Tomasevic, N. (2007). The role of visual and nonvisual feedback in a vehicle steering task. *Journal of Experimental Psychology: Human Perception and Performance, 33*(5), 1127–1144.

Warren, W. H., & Kurz, K. J. (1992). The role of central and peripheral vision in perceiving the direction of self-motion. *Perception & Psychophysics, 51*(5), 443–454.

Warren, W. H., Mestre, D. R., Blackwell, A. W., & Morris, M. W. (1991). Perception of circular heading from optical flow. *Journal of Experimental Psychology: Human Perception and Performance., 17*(1), 28–43.

Weichert, C. (2005). *Möglichkeiten und Grenzen der Blickbewegungsmessung bei Realfahrten [Possibilities and limitations of eye-movement measurement in real driving].* Unpublished student paper, TU Dresden, Chair of Road Planning and Road Design, Dresden.

Weir, D. H., & Chao, K. C. (2007). Control theory models of driving behaviour. In C. Cacciabue (Ed.), *Modelling driver behaviour in automotive environments* (pp. 293–311). London: Springer.

Weise, G., & Durth, W. (1997). *Straßenbau, Planung und Entwurf [Road construction, planning and design]*. Berlin: Verlag für Bauwesen.

Weiße, B. (2005). *Untersuchung des Zusammenhangs von Blickverhaltensparametern und Workload beim Einsatz eines Fahrerassistenzsystems in einem Fahrsimulatorexperiment [Analysis of the relationship between gaze-behaviour parameters and workload when using a driver assistance system in a driving simulator experiment].* Unpublished diploma thesis, TU Dresden, Chair of Traffic & Transportation Psychology.

Weller, G., & Geertsema, K. (2008). Werden ältere Fahrer durch die Fahraufgabe stärker beansprucht als jüngere? [Does the driving task increase older drivers' workload more than younger drivers'? In B. Schlag (Ed.), *Leistungsfähigkeit und Mobilität im Alter [Capability and mobility in old age]* (Vol. 3, 'Mobilität und Alter' [Mobility and age]; Eugen-Otto-Butz Stiftung, pp. 85–111). Köln: TÜV Media.

Weller, G., & Schlag, B. (2004). Verhaltensadaptation nach Einführung von Fahrerassistenzsystemen [Behavioural adaptation following the introduction of driver assistance systems]. In B. Schlag (Ed.), *Verkehrspsychologie. Mobilität – Verkehrssicherheit – Fahrerassistenz [Traffic psychology. Mobility - safety - driving assistance]* (pp. 351–370). Lengerich: Pabst Science Publ.

Weller, G., & Schlag, B. (2007). *RiPCORD-iSEREST. Deliverable D8. Road user behaviour model. (EC: Contract no.: 50 61 84, 6th Framework Programme)*. Retrieved July 13, 2008, from http://ripcord.bast.de/pdf/RIPCORD-ISEREST-Deliverable-D8-Final.pdf

Weller, G., Schlag, B., Friedel, T., & Rammin, C. (2008). Behaviourally relevant road categorisation: A step towards self-explaining rural roads. *Accident Analysis & Prevention, 40*(4), 1581–1588.

Weller, G., Schlag, B., Gatti, G., Jorna, R., & van de Leur, M. (2006). *RiPCORD – iSEREST. Internal report D8.1. Human factors in road design. State of the art and empirical evidence (EC: Contract no.: 50 61 84, 6th Framework Programme)*. Retrieved July 13, 2008, from http://ripcord.bast.de/pdf/RI-TUD-WP8-R1-Human_Factors.pdf

Weller, G., Schlag, B., & Nirschl, G. (2006). *Objective evaluation of ADAS impact on traffic safety by using psychological parameters and driving errors.* Paper presented at the 26th International Congress of Applied Psychology, 16–21 July 2006, Athens, Greece.

Weller, G., Weiße, B., & Schlag, B. (2008). Entwicklung und erste Validierung eines Blickdichteindexes zur Bewertung von Fahrerinformations- und Fahrerassistenzsystemen [Development and first validation of a gaze-density index for the evaluation of driver information and driver assistance systems]. In J. Schade & A. Engeln (Eds.), *Fortschritte der Verkehrspsychologie [Advances in traffic psychology]* (pp. 215–235). Wiesbaden: Verlag für Sozialwissenschaften.

Wendsche, J., Uhmann, S., & Meier, J. (2006). *Beanspruchungsmessung in Kurven: Untersuchung physiologischer und fahrverhaltensbezogener Parameter in Kurven mit unterschiedlicher Unfallhäufigkeit [Measuring workload in curves: A study of physiological and driving-behaviour parameters in curves of different accident frequencies].* Unpublished student paper, TU Dresden, Chair of Traffic and Transportation Psychology, Dresden.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 63–102). London: Academic Press.

Wickens, C. D. (1991). Processing resources and attention. In D. L. Damos (Ed.), *Multiple-task performance*. London: Taylor & Francis.

Wickens, C. D. (1992). *Engineering psychology and human performance* (2nd ed.). New York: Harper-Collins.

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors, 50*(3), 449–455.

Wierwille, W. W., & Eggemeier, F. T. (1993). Recommendations for mental workload measurement in a test and evaluation environment. *Human Factors, 35*(2), 263–281.

Wikman, A.-S., Nieminen, T., & Summala, H. (1998). Driving experience and time-sharing during in-car tasks on roads of different width. *Ergonomics, 41*(3), 358–372.

Wilde, G. J. S. (1988). Risk homeostasis theory and traffic accidents: Propositions, deductions and discussions of dissension in recent reactions. *Ergonomics, 31*(4), 441–468.

Wilde, G. J. S. (1994). *Target risk*. Toronto: PDE.

Wilde, G. J. S. (2001). *Target risk 2. A new psychology of safety and health; what works? What doesn't? And why ...* Toronto: PDE.

Wilde, G. J. S., Claxton-Oldfield, S. P., & Platenius, P. H. (1985). Risk homeostasis in an experimental context. In L. Evans & R. C. Schwing (Eds.), *Human behavior and traffic safety* (pp. 119–149). New York: Plenum Press.

Wilson, J. R., & Rutherford, A. (1989). Mental models: Theory and application in human factors. *Human Factors, 31*(6), 617–634.

Wundt, W. (1874). *Grundzüge der physiologischen Psychologie [Principles of physiological psychology]*. Leipzig: Engelmann; available from http://vlp.mpiwg-berlin.mpg.de/library/data/lit46.

Yan, X., Abdel-Aty, M., & Radwan, E. (2008). Validating a driving simulator using surrogate safety measures. *Accident Analysis & Prevention, 40*(1), 274–288.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology, 18*, 459–482.

Yilmaz, E. H., & Warren, W. H. (1995). Visual control of braking: A test of the tau-dot hypothesis. *Journal of Experimental Psychology: Human Perception and Performance, 21*(5), 996–1014.

York, I., Bradbury, A., Reid, S., Ewings, T., & Paradise, R. (2007). *The manual for streets: Evidence and research. TRL report TRL661*. Retrieved July 3, 2008, from http://www.dft.gov.uk/pgr/sustainable/manforstreets/manualforstreetsevidence

Young, M. S., & Stanton, N. A. (2002). Malleable attentional resources theory: A new explanation for the effects of mental underload on performance. *Human Factors, 44*(3), 365–375.

Zeitlin, L. R. (1995). Estimates of driver mental workload: A long-term field trial of two subsidiary tasks. *Human Factors, 37*(3), 611–621.

Zeitlin, L. R. (1998). Micromodel for objective estimation of driver mental workload from task data. *Transportation Research Record, 1631*, 28–34.

Zimbardo, P. G., & Gerrig, R. J. (2004). *Psychologie [Psychology]*. München: Pearson Studium.

Zuckerman, M. (1976). Sensation seeking and anxiety, traits and states, as determinants of behavior in novel situations. In I. G. Sarason & C. D. Spielberger (Eds.), *Stress and anxiety* (pp. 141–170). Oxford, England: Hemisphere.

Zuckerman, M. (1994). *Behavioral expressions and biosocial bases of sensation seeking*. New York: Cambridge University Press.

Zuckerman, M. (2007). *Sensation seeking and risky behavior*. Washington, DC: American Psychological Association.

Zwielich, F., Reker, K., & Flach, J. (2001). Fahrerverhaltensbeobachtungen auf Landstraßen am Beispiel von Baumalleen - eine Untersuchung mit dem Fahrzeug zur Interaktionsforschung Straßenverkehr [Surveillance of driver behaviour on country roads taking the example of tree-lined avenues (translation taken from the source)]. In Bundesanstalt für Straßenwesen (Ed.), *Berichte der Bundesanstalt für Straßenwesen. Heft M124*. Bremerhaven: Wirtschaftsverlag NW.

Appendix

Contents Appendix

To access the book's appendix, please visit our OnlinePLUS program at www.vs-verlag.de and search for the author's name.